



HAL
open science

Anisotropic multivariate deconvolution using projection on the Laguerre basis

Florian Dussap

► **To cite this version:**

Florian Dussap. Anisotropic multivariate deconvolution using projection on the Laguerre basis. 2020.
hal-02400684v2

HAL Id: hal-02400684

<https://hal.science/hal-02400684v2>

Preprint submitted on 6 Jan 2020 (v2), last revised 10 Jul 2020 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Anisotropic multivariate deconvolution using projection on the Laguerre basis

Florian Dussap

Université de Paris, MAP5, UMR CNRS 8145, F-75006 Paris, France

Abstract

We investigate adaptive density estimation in the additive model $Z = X + Y$, where X and Y are independent d -dimensional random vectors with non-negative coordinates. Our goal is to recover the density of X from independent observations of Z , assuming the density of Y is known. In the $d = 1$ case, an estimation procedure using projection on the Laguerre basis have already been studied. We generalize this procedure in the multivariate case: we establish non-asymptotic upper bounds on the mean integrated squared error of the estimator and we derive convergence rates on anisotropic functional spaces. Moreover, we provide data-driven strategies for selecting the right projection space (for $d = 1$, we improve the previous projection procedure). We illustrate these procedures on simulated data, and in dimension $d = 1$ we compare our procedure with the previous adaptive projection procedure.

Keywords: anisotropic multivariate projection estimator, laguerre basis, model selection, nonparametric density estimation
2010 MSC: 62G07, 62H12

1. Introduction

1.1. Statistical model and motivations

In this article, we study the problem of recovering the distribution of a random vector X when we only observe its sum with a noise vector Y with known distribution. This is a classical problem in nonparametric statistics (see references below), but we focus on the particular case where both X and Y have non-negative coordinates. This assumption is quite unusual in deconvolution problems but is relevant for instance in survival analysis.

More precisely, we consider the following statistical model:

$$Z_i = X_i + Y_i, \quad i = 1, \dots, n \quad (1)$$

where the X_i 's, the Y_i 's and Z_i 's are random vectors in \mathbb{R}^d with non-negative coordinates. We assume that the X_i 's are i.i.d. with unknown density f on \mathbb{R}_+^d , and that the Y_i 's are i.i.d. with known density g on \mathbb{R}_+^d . Moreover, we assume that the X_i 's and the Y_i 's are independent. Our goal is to provide an adaptive procedure to estimate the density f from the observations Z_1, \dots, Z_n .

In the univariate case, there are a lot of papers about recovering the density of a random variable when it is observed with an additive known noise. Many authors use a kernel estimator

Email address: florian.dussap@parisdescartes.fr (Florian Dussap)

introduced by Stefanski and Carroll (1990). Fan (1991) first introduced the notion of *ordinary smooth* and *supersmooth* noise (rate of decay of the characteristic function) to study optimal rates of convergence on Hölder spaces. When f is supersmooth and the noise is ordinary smooth, Butucea (2004) showed that the kernel estimator achieves a good rate of convergence and that this rate is optimal. When both f and the noise are supersmooth, the problem is more complicated and have been studied by Butucea and Tsybakov (2008a,b) from lower bound point of view.

To provide an adaptive estimator, different procedures were proposed. When f belongs to a Sobolev space, Pensky and Vidakovic (1999) proposed a wavelet strategy that is adaptive and achieves optimal rates of convergence. For kernel estimators, Delaigle and Gijbels (2004) estimated the optimal bandwidth with a bootstrap procedure and showed its consistency. Hazelton and Turlach (2009, 2010) proposed a weighted kernel estimator with a data-driven way to choose the weights. Moreover, their weighted kernel estimator can be used in a multivariate setting. For projection with penalization strategies, Comte et al. (2006) used a Shannon type basis to construct an adaptive estimator that is minimax in most cases. More recently, non-compact supported bases were used by Mabon (2017) (Laguerre basis), by Comte and Genon-Catalot (2018) (Laguerre basis and Hermite basis), and by Sacko (2019) (Hermite basis) to construct adaptive estimators on suitable functional spaces.

The multivariate deconvolution literature is more sparse. Masry (1991) generalizes the kernel estimator for stationary random processes, with a dependence structure between the variables. The noise is assumed to have i.i.d. coordinates (isotropic noise) and no adaptive strategy is proposed, the author focuses on the problem of dependency between the variables. Youndjé and Wells (2008) propose a cross-validation strategy to estimate the optimal bandwidth of the kernel estimator in the multivariate setting, and show it is asymptotically optimal under the assumption that the noise is isotropic and ordinary smooth. Comte and Lacour (2013) use a bandwidth selection procedure inspired by Goldenshluger and Lepski (2011). Their procedure allows anisotropic noises, with both ordinary smooth and supersmooth components, and they derive rates of convergence for the pointwise risk and the L^2 risk, when f belongs to anisotropic Hölder, Nikol'skii or Sobolev classes. For ordinary smooth noise and when f belongs to anisotropic Nikol'skii classes, Rebelles (2016) provides an adaptive kernel estimator which is minimax for the L^p -loss. Recently, Lepski and Willer (2019) studied a more general model (with direct and indirect observations of X) and provided an adaptive kernel estimator on anisotropic Nikol'skii classes, under the L^p -loss.

Concerning our specific case of deconvolution with non-negative noise, the case $d = 1$ has already been studied by Mabon (2017) using a projection strategy on the Laguerre basis. We also use a projection strategy in the multivariate case. The main tool we use to construct our estimator is the theory of hypermatrices. Using the contraction product of hypermatrices, we show it is possible to recover the coefficients of f from the observations. We recall the definitions of these objects in Subsection 1.2.

We provide rates of convergence for the MISE of our estimator on anisotropic functional spaces: Sobolev–Laguerre spaces and smooth Laguerre spaces. We propose a model selection procedure to produce an adaptive estimator, under mild assumptions on the noise density g . This procedure is inspired by the work of Goldenshluger and Lepski (2011) concerning bandwidth selection. It was introduced for model selection by Chagny (2013b) for estimation of conditional density, in a bi-dimensional setting. We show this procedure can be applied to our deconvolution problem, in a d -dimensional setting, and we establish an asymptotic oracle inequality for this procedure. Moreover, the proof is written to provide general steps that can be applied to other contexts.

Outline of the paper. In Section 2, we construct the estimator and we study its MISE. We provide non asymptotic MISE bounds and we derive convergence rates on Sobolev–Laguerre balls and smooth Laguerre balls. In Section 3, we give model selection procedures to construct an adaptive estimator and we establish oracle inequalities. In Section 4, we illustrate the procedures on simulated data. All the proofs are gathered in Section 5.

In the remaining of this section, we give the notations and the key facts about hypermatrices we will use in the rest of the paper.

1.2. Preliminary on hypermatrices

Let $\alpha, \beta \in \mathbb{N}^d$ be multi-indices and $x \in \mathbb{R}^d$. We will use the following notations:

$$\begin{aligned} \alpha + \beta &:= (\alpha_1 + \beta_1, \dots, \alpha_d + \beta_d) & \alpha \times \beta &:= (\alpha_1 \beta_1, \dots, \alpha_d \beta_d) \\ \frac{\alpha}{\beta} &:= \left(\frac{\alpha_1}{\beta_1}, \dots, \frac{\alpha_d}{\beta_d} \right) & x^\alpha &:= x_1^{\alpha_1} \dots x_d^{\alpha_d} \\ |\alpha| &:= \alpha_1 + \dots + \alpha_d & \beta \leq \alpha &\iff \forall q \in \{1, \dots, d\}, \beta_q \leq \alpha_q \\ m \wedge m' &:= (m_1 \wedge m'_1, \dots, m_d \wedge m'_d) \text{ where } m_i \wedge m'_i := \min(m_i, m'_i). \end{aligned}$$

We denote $\mathbf{0}$ the multi-index $(0, \dots, 0)$ and $\mathbf{1}$ the multi-index $(1, \dots, 1)$. If $k \in \mathbb{N}^d$ and $\ell \in \mathbb{N}^p$ are two multi-indices, we denote $(k, \ell) \in \mathbb{N}^{d+p}$ their concatenation.

Proposition 1.1 (Multinomial theorem). *If $\alpha \in \mathbb{N}^d$ is a multi-index and if $x, y \in \mathbb{R}^d$ are vectors, then $(x + y)^\alpha = \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} x^\beta y^{\alpha - \beta}$ where $\binom{\alpha}{\beta} := \binom{\alpha_1}{\beta_1} \times \dots \times \binom{\alpha_d}{\beta_d}$.*

We recall some facts about hypermatrices; see (Hogben, 2013) for more details. A hypermatrix is a multi-dimensional array of numbers. The order of a hypermatrix is its number of dimensions : an order-1 hypermatrix is a vector, an order-2 hypermatrix is matrix, and so on. If $m = (m_1, \dots, m_d) \in (\mathbb{N}^*)^d$ is multi-index, we denote by $\mathbb{R}^m = \mathbb{R}^{(m_1, \dots, m_d)}$ the set of order- d hypermatrices with m_j components according to the j -th dimension. Moreover, if $m \in (\mathbb{N}^*)^d$ and $r \in (\mathbb{N}^*)^p$ are two multi-indices, since (m, r) denotes the concatenation of m and r , we denote $\mathbb{R}^{(m, r)} := \mathbb{R}^{(m_1, \dots, m_d, r_1, \dots, r_p)}$. We choose to index the hypermatrix components starting from 0. So if $T \in \mathbb{R}^m$ is a hypermatrix, its components are T_k for $k \in \mathbb{N}^d$ such that $k \leq m - \mathbf{1}$.

Definition 1.2. Let $m \in (\mathbb{N}^*)^d$, we make \mathbb{R}^m an euclidean vector space, defining the addition and the scalar multiplication by $[T + U]_k := T_k + U_k$ and $[\lambda T]_k := \lambda T_k$, and defining the scalar product and the associated norm by $\langle T, U \rangle_{\mathbb{R}^m} := \sum_{k \leq m - \mathbf{1}} T_k U_k$ and $\|T\|_{\mathbb{R}^m}^2 := \sum_{k \leq m - \mathbf{1}} T_k^2$. For vectors ($d = 1$), we get the usual ℓ^2 norm and for matrices ($d = 2$), we get the Frobenius norm.

Definition 1.3 (Contraction product). Let m, p et r be multi-indices (with possibly different lengths) and denote d the length of p . Let $T \in \mathbb{R}^{(m, p)}$ and $U \in \mathbb{R}^{(p, r)}$ be two hypermatrices, their d -contraction product is the hypermatrix $T \times_d U \in \mathbb{R}^{(m, r)}$ defined by:

$$[T \times_d U]_{(k, \ell)} := \sum_{j \leq p - \mathbf{1}} T_{kj} U_{j\ell}. \quad (2)$$

The contraction product generalizes the matrix product: if T , respectively U , are matrices $m \times p$, respectively $p \times r$, their 1-contraction product is their matrix product.

Definition 1.4 (Frobenius norm and spectral norm). Let $m \in (\mathbb{N}^*)^d$ be a multi-index. If $T \in \mathbb{R}^{(m, m)}$ is a hypermatrix, then T induces a linear map (still denoted by T):

$$\begin{aligned} T: \mathbb{R}^m &\longrightarrow \mathbb{R}^m \\ x &\longmapsto T \times_d x \end{aligned}$$

We denote $T^* \in \mathbb{R}^{(m,m)}$ the hypermatrix defined by $[T^*]_{k\ell} = T_{\ell k}$. The linear map induced by T^* is the adjoint linear map of the one induced by T :

$$\forall x, y \in \mathbb{R}^m, \quad \langle y, Tx \rangle_{\mathbb{R}^m} = \langle T^*y, x \rangle_{\mathbb{R}^m}.$$

We define the Frobenius norm and the spectral norm of T in the same way we define them for any endomorphism: $\|T\|_F^2 := \text{Tr}(T^*T)$ and $\rho^2(T) := \lambda_{\max}(T^*T)$.

It is easy to check that:

$$\|T\|_F^2 = \sum_{k, \ell \leq m-1} T_{k\ell}^2, \quad \rho(T) = \sup_{x \in \mathbb{R}^m \setminus \{0\}} \frac{\|Tx\|_{\mathbb{R}^m}}{\|x\|_{\mathbb{R}^m}}.$$

Lastly, we recall the classical inequalities between the Frobenius and the spectral norm:

$$\frac{1}{m_1 \cdots m_d} \|T\|_F^2 \leq \rho^2(T) \leq \|T\|_F^2. \quad (3)$$

2. Study of the estimator

2.1. Estimation procedure

In the model (1), the Z_i 's are i.i.d. random vectors on \mathbb{R}_+^d , and they admit a density we denote by h . This density function is given by the convolution product of f and g :

$$(f * g)(x) := \int_{\mathbb{R}^d} f(u) g(x - u) du = \int_{[0, x_1] \times \cdots \times [0, x_d]} \cdots \int f(u) g(x - u) du.$$

We assume that f, g and h belong to $L^2(\mathbb{R}_+^d)$ and we expand them in the multivariate Laguerre basis. We recall that the one-dimensional Laguerre functions $(\varphi_k)_{k \in \mathbb{N}}$ are defined by:

$$\forall x \in \mathbb{R}_+, \quad \varphi_k(x) := \sqrt{2} L_k(2x) e^{-x}, \quad \text{where } L_k(x) := \sum_{j=0}^k \binom{k}{j} \frac{(-x)^j}{j!},$$

and that they form an orthonormal basis of $L^2(\mathbb{R}_+)$. In the multivariate case, we *tensorize* the Laguerre basis. For $k = (k_1, \dots, k_d) \in \mathbb{N}^d$ a multi-index, we define the multivariate Laguerre function φ_k on \mathbb{R}_+^d as the tensor product of one-dimensional Laguerre functions:

$$\varphi_k(x_1, \dots, x_d) := (\varphi_{k_1} \otimes \cdots \otimes \varphi_{k_d})(x_1, \dots, x_d) = \varphi_{k_1}(x_1) \times \cdots \times \varphi_{k_d}(x_d). \quad (4)$$

The multivariate Laguerre functions $(\varphi_k)_{k \in \mathbb{N}^d}$ forms a basis of $L^2(\mathbb{R}_+^d)$, and we expand the functions f, g and h in this basis:

$$f = \sum_{k \in \mathbb{N}^d} a_k \varphi_k, \quad g = \sum_{j \in \mathbb{N}^d} b_j \varphi_j, \quad h = \sum_{\ell \in \mathbb{N}^d} c_\ell \varphi_\ell. \quad (5)$$

The use of the Laguerre basis is relevant in this context, as the one-dimensional Laguerre functions verify the relation:

$$\forall k, j \in \mathbb{N}, \quad \varphi_k * \varphi_j = 2^{-1/2} (\varphi_{k+j} - \varphi_{k+j+1}), \quad (6)$$

see (Abramowitz and Stegun, 1972, formula 22.13.14). Using this relation and $h = f * g$, by expanding the functions f, g and h on the Laguerre basis, we get a relation between their coefficients.

Proposition 2.1. *If a , b and c are the coefficients defined in (5), then the following relation holds:*

$$\forall \ell \in \mathbb{N}^d, \quad c_\ell = 2^{-d/2} \sum_{\varepsilon \in \{0,1\}^d} (-1)^{|\varepsilon|} (a * b)_{\ell - \varepsilon}$$

where $a * b$ is the discrete convolution product of a and b defined by $(a * b)_k := \sum_{j \leq k} a_j b_{k-j}$ if $k \in \mathbb{N}^d$ and $(a * b)_k = 0$ if $k \notin \mathbb{N}^d$.

This relation can be written as a discrete convolution product $c = \beta * a$ with $\beta \in \mathbb{N}^d$ defined by:

$$\beta_k := 2^{-d/2} \sum_{\varepsilon \in \{0,1\}^d} (-1)^{|\varepsilon|} b_{k - \varepsilon} \quad (7)$$

where by convention $b_j = 0$ if $j \notin \mathbb{N}^d$. Thus, we have a linear relation between the coefficients c and a :

$$\forall \ell \in \mathbb{N}^d, \quad c_\ell = \sum_{k \leq \ell} \mathbf{G}_{\ell k} a_k, \quad \text{where } \mathbf{G}_{\ell k} := \begin{cases} \beta_{\ell - k} & \text{if } k \leq \ell \\ 0 & \text{else.} \end{cases} \quad (8)$$

If we consider \mathbf{G} as an infinite hypermatrix $[\mathbf{G}_{\ell k}]_{\ell, k \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d \times \mathbb{N}^d}$, then it is *lower triangular* according to the following definition.

Definition 2.2. A hypermatrix $T \in \mathbb{R}^{(m,m)}$ is said to be lower triangular if apart from multi-indices $\ell, k \leq m - \mathbf{1}$ such that $k \leq \ell$, we have $T_{\ell k} = 0$.

An infinite hypermatrix $T \in \mathbb{R}^{\mathbb{N}^d \times \mathbb{N}^d}$ is said to be lower triangular if apart from multi-indices $\ell, k \in \mathbb{N}^d$ such that $k \leq \ell$, we have $T_{\ell k} = 0$.

In the next proposition, we show that the linear relation between a and c is invertible.

Proposition 2.3. *Let \mathbf{G} be the infinite hypermatrix defined in (8). For every $k \in \mathbb{N}^d$, there exists $[\mathbf{H}_{k\ell}]_{\ell \leq k}$ such that for every $a, c \in \mathbb{N}^d$ satisfying the relation (8), we have:*

$$a_\ell = \sum_{\ell \leq k} \mathbf{H}_{\ell k} c_k. \quad (9)$$

We denote $\mathbf{H}_{\ell k} := (\mathbf{G}^{-1})_{\ell k}$.

We can write the linear relations between a and c using hypermatrices and contraction products. For $m \in (\mathbb{N}^*)^d$, we denote by \mathbf{a}_m (resp. \mathbf{c}_m) the hypermatrix $[a_k]_{k \leq m - \mathbf{1}} \in \mathbb{R}^m$ (resp. $[c_\ell]_{\ell \leq m - \mathbf{1}} \in \mathbb{R}^m$), and we denote by \mathbf{G}_m and \mathbf{G}_m^{-1} the hypermatrices $[\mathbf{G}_{\ell k}]_{\ell, k \leq m - \mathbf{1}}$ and $[(\mathbf{G}^{-1})_{\ell k}]_{\ell, k \leq m - \mathbf{1}}$ in $\mathbb{R}^{(m,m)}$. Then, we have:

$$\mathbf{c}_m = \mathbf{G}_m \times_d \mathbf{a}_m \quad \mathbf{a}_m = \mathbf{G}_m^{-1} \times_d \mathbf{c}_m, \quad (10)$$

where “ \times_d ” stands for the contraction product defined by (2).

Estimation procedure. For $m \in (\mathbb{N}^*)^d$, let S_m be the vector space spanned by the functions φ_k for $k \leq m - \mathbf{1}$, and let $D_m := m_1 \cdots m_d$ be its dimension. We estimate f by estimating f_m the projection of f on S_m . This projection is given by $f_m = \sum_{k \leq m - \mathbf{1}} a_k \varphi_k$, so the problem reduces to the estimation of \mathbf{a}_m . Because \mathbf{a}_m is related to \mathbf{c}_m by (10) and since $c_\ell = \mathbb{E}[\varphi_\ell(Z_1)]$, we estimate f by:

$$\hat{f}_m := \sum_{k \leq m - \mathbf{1}} \hat{a}_k \varphi_k \quad \text{where } \hat{\mathbf{a}}_m := \mathbf{G}_m^{-1} \times_d \hat{\mathbf{c}}_m \quad \text{and } \hat{c}_\ell := \frac{1}{n} \sum_{i=1}^n \varphi_\ell(Z_i).$$

2.2. Non-asymptotic MISE bounds

We quantify the quality of the estimator \hat{f}_m by its MISE (Mean Integrated Squared Error): $\mathbb{E}\|f - \hat{f}_m\|_{L^2}^2$. In the next proposition, we decompose the MISE in a bias term and a variance term and give a bound on the MISE of \hat{f}_m .

Proposition 2.4. *If f and g are $L^2(\mathbb{R}_+^d)$ functions, then we have the inequality:*

$$\forall m \in \mathbb{N}^d, \quad \mathbb{E}\|f - \hat{f}_m\|_{L^2}^2 \leq \|f - f_m\|_{L^2}^2 + \frac{2^d D_m \rho^2(\mathbf{G}_m^{-1})}{n} \wedge \frac{\|h\|_\infty \|\mathbf{G}_m^{-1}\|_F^2}{n}.$$

Remark 2.5. Using the Cauchy–Schwarz inequality, it holds $\|h\|_\infty \leq \|f\|_{L^2} \|g\|_{L^2}$ which is finite by assumption. Moreover, if g is bounded, we also have $\|h\|_\infty \leq \|g\|_\infty$.

Remark 2.6. The norm equivalence (3) implies that $\|\mathbf{G}_m^{-1}\|_F^2 \leq D_m \rho^2(\mathbf{G}_m^{-1})$, so the order of magnitude of the variance term is given by $\|\mathbf{G}_m^{-1}\|_F^2$. The minimum is important only for the small values of m (because of the constants).

Under an additional assumption, Comte and Genon-Catalot (2018) improved the variance bound in the one-dimensional case. We generalize their result to the multivariate case.

Proposition 2.7. *We denote $Y_1^{(j)}$ the j -th coordinate of Y_1 . We assume that for any nonempty subset J of $\{1, \dots, d\}$, we have:*

$$M_J(g) := \mathbb{E} \left[\prod_{j \in J} \frac{1}{\sqrt{Y_1^{(j)}}} \right] < +\infty. \quad (11)$$

Then we have the following inequality:

$$\forall m \in (\mathbb{N}^*)^d, \quad \mathbb{E}\|f - \hat{f}_m\|_{L^2}^2 \leq \|f - f_m\|_{L^2}^2 + \frac{c(g) \sqrt{D_m} \rho^2(\mathbf{G}_m^{-1})}{n} \wedge \frac{\|h\|_\infty \|\mathbf{G}_m^{-1}\|_F^2}{n},$$

where $c(g)$ is a positive constant depending on $M_J(g)$ for $J \subseteq \{1, \dots, d\}$.

2.3. Convergence rates for the MISE

To study the bias term, we assume that f belongs to a Sobolev–Laguerre space. In dimension $d = 1$, these functional spaces have been introduced by Bongioanni and Torrea (2009) to study the Laguerre operator. The connection with Laguerre coefficients was established later by Comte and Genon-Catalot (2015). Following the same idea, we define Sobolev–Laguerre balls on \mathbb{R}_+^d .

Definition 2.8 (Sobolev–Laguerre ball). Let $L > 0$ and $s \in (0, +\infty)^d$, we define the Sobolev–Laguerre ball of order s and radius L by:

$$W^s(\mathbb{R}_+^d, L) := \left\{ f \in L^2(\mathbb{R}_+^d) \left| \sum_{k \in \mathbb{N}^d} a_k^2(f) k^s \leq L \right. \right\},$$

with $a_k(f) := \langle f, \varphi_k \rangle_{L^2}$ the Laguerre coefficients of f .

Assuming f to belong to $W^s(\mathbb{R}_+^d, L)$, the bias term decreases to 0 with polynomial rate. Indeed, for $m \in (\mathbb{N}^*)^d$, we have:

$$\|f - f_m\|_{L^2}^2 = \sum_{\substack{k \in \mathbb{N}^d \\ \exists q, k_q \geq m_q}} a_k^2(f) \leq \sum_{q=1}^d \sum_{\substack{k \in \mathbb{N}^d \\ k_q \geq m_q}} a_k^2(f) k_q^{s_q} k_q^{-s_q} \leq L \sum_{q=1}^d m_q^{-s_q}. \quad (12)$$

The case where the Laguerre coefficients of f decrease with exponential rate is also interesting. We define new functional spaces, “*smooth Laguerre spaces*”, in the following way.

Definition 2.9 (Smooth Laguerre ball). Let $L > 0$ and $r \in (0, +\infty)^d$, we define the smooth Laguerre ball of order r and radius L as:

$$\mathcal{S}^r(\mathbb{R}_+^d, L) := \left\{ f \in L^2(\mathbb{R}_+^d) \left| \sum_{k \in \mathbb{N}^d} a_k^2(f) e^{r \cdot k} \leq L \right. \right\}.$$

By the same argument as previously, if f belongs to $\mathcal{S}^r(\mathbb{R}_+^d, L)$, the bias term decreases to 0 with exponential rate:

$$\forall m \in (\mathbb{N}^*)^d, \quad \|f - f_m\|_{L^2}^2 \leq L \sum_{q=1}^d e^{-r_q m_q}. \quad (13)$$

Now, we need to control the variance term in Proposition 2.4. In the one-dimensional case, this control can be provided under assumptions on the Laplace transform G of g and under assumptions on the derivatives of g , see lemma 3.6 in (Comte et al., 2017). In the next proposition, we extend this result to the multivariate case. Moreover, we make assumptions only on the behavior of the Laplace transform of g , we do not need to study its differentials.

We recall that the Laplace transform of g is the function G defined on the domain \mathcal{P}_+^d by:

$$G(s) := \int_{\mathbb{R}_+^d} e^{-s \cdot x} g(x) dx,$$

where \mathcal{P}_+ stands for the set of complex numbers with non-negative real part. In addition, we extend the set \mathbb{C} of complex numbers by adding a point at infinity denoted by ∞ . The control of the Frobenius norm of \mathbf{G}_m^{-1} relies on the behavior of G when some of its arguments take the ∞ value.

Proposition 2.10. *We assume that $\beta \in \ell^1(\mathbb{N}^d)$, with β defined in (7). We assume that G is non-zero on \mathcal{P}_+^d and that there exists $\alpha \in (\mathbb{N}^*)^d$ such that the function:*

$$K_\alpha(s) := (1 + s)^\alpha G(s), \quad s \in \mathcal{P}_+^d$$

can be extended as a non-zero function on $(\mathcal{P}_+ \cup \{\infty\})^d$ such that the restriction of K_α on $(i\mathbb{R} \cup \{\infty\})^d$ is continuous. Then for $m \in \mathbb{N}^d$ satisfying $m \geq 4$, there exists a constant $C > 0$ depending on β such that $\|\mathbf{G}_m^{-1}\|_F^2 \leq Cm^{2\alpha}$.

Remark 2.11. If $g \in W^s(\mathbb{R}_+^d, L)$ with $L > 0$ and $s \in (1, +\infty)^d$, the Laguerre coefficients of g belong to $\ell^1(\mathbb{N}^d)$. Indeed, using Cauchy–Schwarz inequality,

$$\sum_{k \in \mathbb{N}^d} |b_k| = \sum_{k \in \mathbb{N}^d} |b_k| k^{\frac{s}{2}} k^{-\frac{s}{2}} \leq \left(\sum_{k \in \mathbb{N}^d} b_k^2 k^s \right)^{\frac{1}{2}} \left(\sum_{k \in \mathbb{N}^d} k^{-s} \right)^{\frac{1}{2}} < +\infty$$

because for every $q \in \{1, \dots, d\}$, $s_q > 1$.

Remark 2.12. We notice that in dimension $d = 1$, these assumptions simply become:

1. The Laplace transform G does not vanish on \mathcal{P}_+ .
2. The Fourier transform of g admits an asymptotic expansion $g^*(\omega) = \omega^{-\alpha} (K_\alpha + o(1))$ when $|\omega|$ goes to $+\infty$, for some $\alpha \in \mathbb{N}^*$ and some non-zero constant K_α .

It is easy to see that this second assumption is a consequence of the assumptions on the derivatives of g made in (Comte et al., 2017, subsection 2.5).

Remark 2.13. If the distribution of Y is a product of gamma distributions $\bigotimes_{q=1}^d \Gamma(\alpha_q, \lambda_q)$ with $\alpha \in (\mathbb{N}^*)^d$, then G and K_α are given by:

$$G(s) = \prod_{q=1}^d \left(1 + \frac{s_q}{\lambda_q}\right)^{-\alpha_q} \quad \text{and} \quad K_\alpha(s) = \lambda^\alpha \prod_{q=1}^d \left(\frac{1 + s_q}{\lambda_q + s_q}\right)^{\alpha_q},$$

so the assumptions of Proposition 2.10 are fulfilled.

Gathering (12) or (13) with Proposition 2.10, we obtain convergence rates for our estimator.

Theorem 2.14. *Let $s, r \in (0, +\infty)^d$ and $L > 0$. Assume that g satisfies the assumptions of Proposition 2.10 with $\alpha \in (\mathbb{N}^*)^d$.*

1. *For $m_{\text{opt}} \in (\mathbb{N}^*)^d$ given by $m_{\text{opt},j} \propto n^{1/(s_j + s_j \sum_{i=1}^d \frac{2\alpha_i}{s_i})}$, there exists a constant $C > 0$ depending on s, L and g such that:*

$$\sup_{f \in \mathcal{W}^s(\mathbb{R}_+^d, L)} \mathbb{E} \|f - \hat{f}_{m_{\text{opt}}}\|_{L^2}^2 \leq C n^{-1/(1 + \sum_{i=1}^d \frac{2\alpha_i}{s_i})}.$$

2. *For $m_{\text{opt}} \in (\mathbb{N}^*)^d$ given by $m_{\text{opt},j} \propto \frac{\log n}{r_j}$, there exists a constant $C > 0$ depending on r, L and g such that:*

$$\sup_{f \in \mathcal{S}^r(\mathbb{R}_+^d, L)} \mathbb{E} \|f - \hat{f}_{m_{\text{opt}}}\|_{L^2}^2 \leq C \frac{(\log n) \sum_{i=1}^d 2\alpha_i}{n}.$$

Remark 2.15. Our convergence rates on Sobolev–Laguerre balls are similar to convergence rates found by Comte and Lacour (2013) on anisotropic Sobolev balls, in the context of deconvolution using a kernel estimator with an ordinary smooth noise.

3. Adaptive estimation

In practice, we do not know the underlying regularity of f , so we can not compute the model m_{opt} of Theorem 2.14. We want a data-driven procedure that automatically provides the bias-variance compromise, without assumptions on f . More precisely, we want to select \hat{m} from the data such that the MISE of $\hat{f}_{\hat{m}}$ is close to the oracle MISE: $\inf_m \mathbb{E} \|f - \hat{f}_m\|_{L^2}^2$.

Let $m^* \in (\mathbb{N}^*)^d$ fixed such that $D_{m^*} \leq n$. We only look at sub-models of m^* that are not “too big”. More precisely, we consider the models that belong to the following model collection:

$$\mathcal{M}_n := \left\{ m \leq m^* \mid D_m \rho^2(\mathbf{G}_m^{-1}) \leq \frac{n}{\log n} \right\}. \quad (14)$$

We assume that the Y_i ’s distribution satisfies the following assumptions.

- (A1) The noise density g is bounded.
- (A2) For every nonempty $J \subseteq \{1, \dots, d\}$, $M_J(g)$ defined by (11) is finite.
- (A3) For every $b > 0$, we have $\sum_{m \in \mathcal{M}_n} \rho^2(\mathbf{G}_m^{-1}) e^{-b\sqrt{D_m}} \leq K(b)$, with $K(b)$ a positive constant not depending on n .

Under assumptions **(A1)** and **(A2)**, we can apply Proposition 2.7 to control the variance term $\mathbb{E}\|\hat{f}_m - f_m\|_{L^2}^2$ by:

$$V(m) := \frac{c(g)\sqrt{D_m}\rho^2(\mathbf{G}_m^{-1})}{n} \wedge \frac{(\|g\|_\infty \vee 1)\|\mathbf{G}_m^{-1}\|_F^2 \log n}{n}. \quad (15)$$

Moreover under assumption **(A3)**, we can control the right deviation of $\|\hat{f}_m - f_m\|_{L^2}^2$ when m belongs to \mathcal{M}_n .

Lemma 3.1. *Under assumptions **(A1)** to **(A3)**, there exists a numerical constant $a_0(d) > 0$ depending on the dimension d such that for every $a > a_0(d)$:*

$$\sum_{m \in \mathcal{M}_n} \mathbb{E} \left[\left(\|\hat{f}_m - f_m\|_{L^2}^2 - aV(m) \right)_+ \right] \leq \frac{C}{n},$$

with C a positive constant depending on g and a .

3.1. Back to the one-dimensional case

We consider two procedures, indexed by $i \in \{1, 2\}$. In both procedures, we choose the model minimizing the penalized criterion:

$$\hat{m}_i := \arg \min_{m \in \mathcal{M}_n} \left[-\|\hat{f}_m\|_{L^2}^2 + \kappa_i \text{pen}_i(m) \right], \quad i \in \{1, 2\} \quad (16)$$

where \mathcal{M}_n is defined by (14), κ_i is a numerical constant to be adjusted and pen_i is the penalty term. Mabon (2017) proposes the following penalty term:

$$\text{pen}_1(m) := \frac{2m\rho^2(\mathbf{G}_m^{-1})}{n} \wedge \frac{(\|g\|_\infty \vee 1)\|\mathbf{G}_m^{-1}\|_F^2 \log n}{n},$$

and provides an oracle inequality for the estimator $\hat{f}_{\hat{m}_1}$ under assumptions **(A1)** and **(A3)**. This choice is based on the bound of the variance term of Proposition 2.4. However if g is bounded, assumption **(A2)** holds automatically in dimension $d = 1$, so we can apply Proposition 2.7 to get a better bound on the variance term. Thus, we propose the penalty term $\text{pen}_2(m) = V(m)$ where $V(m)$ is defined by (15). We show that our estimator $\hat{f}_{\hat{m}_2}$ also satisfies an oracle inequality.

Theorem 3.2. *We assume $d = 1$. Under assumptions **(A1)** to **(A3)**, there exists a numerical constant $\kappa_0 > 0$ such that for every choice of $\kappa > \kappa_0$, we have the following oracle inequality:*

$$\mathbb{E} \left[\|f - \hat{f}_{\hat{m}_2}\|_{L^2}^2 \right] \leq 4 \inf_{m \in \mathcal{M}_n} \left(\|f - f_m\|_{L^2}^2 + \kappa \text{pen}_2(m) \right) + \frac{C}{n},$$

with C a positive constant depending on κ and g .

3.2. Multivariate case

We use the ‘‘Goldenshlugger and Lepski’’ type procedure for model selection introduced by Chagny (2013b) for the estimation of a conditional density function. We apply this procedure to our multivariate deconvolution problem, and we establish an oracle inequality.

We choose \hat{m} in the model collection \mathcal{M}_n defined in (14), minimizing:

$$\hat{m} := \arg \min_{m \in \mathcal{M}_n} [A(m) + \kappa_2 V(m)],$$

where $V(m)$ is defined by (15) and $A(m)$ is a term which has the order of the bias term (see the proof):

$$A(m) := \max_{m' \in \mathcal{M}_n} \left(\|\hat{f}_{m'} - \hat{f}_{m \wedge m'}\|_{L^2}^2 - \kappa_1 V(m') \right)_+,$$

and where κ_1, κ_2 are two numerical constants to be adjusted.

Theorem 3.3 (Oracle inequality). *Under assumptions (A1) to (A3), there exists a numerical constant $\kappa_0(d) > 0$ depending on the dimension d such that for every choice of κ_1, κ_2 satisfying $\kappa_0(d) < \kappa_1 \leq \kappa_2$, we have the following oracle inequality:*

$$\mathbb{E} \left[\|f - \hat{f}_{\hat{m}}\|_{L^2}^2 \right] \leq C \inf_{m \in \mathcal{M}_n} \left(\|f - f_m\|_{L^2}^2 + V(m) \right) + \frac{C'}{n},$$

with C a positive constant depending on κ_1 and κ_2 , and C' a positive constants depending on g , d and κ_1 .

Remark 3.4. It is important to notice that the proof of Theorem 3.3 only depends on the deconvolution structure through Lemma 3.1. The proof relies on the multivariate L^2 structure and on an upper bound $V(m)$ of the variance term $\mathbb{E}\|f_m - \hat{f}_m\|_{L^2}^2$ that satisfies the concentration inequality of Lemma 3.1. Thus, the procedure can be applied to more general model selection problems.

4. Numerical illustrations

4.1. Estimators comparison in dimension 1

We want to compare the two estimators $\hat{f}_{\hat{m}_1}$ and $\hat{f}_{\hat{m}_2}$ defined in Subsection 3.1, with different distributions for X and Y , and with different sample size n . We compare their MISE:

$$\mathcal{R}_i = \mathcal{R}_i(f, g, n) := \mathbb{E} \|f - \hat{f}_{\hat{m}_i}\|_{L^2}^2, \quad i \in \{1, 2\}.$$

We also compute the oracle risk:

$$\mathcal{R}_o = \mathcal{R}_o(f, g, n) := \min_{1 \leq m \leq m^*} \mathbb{E} \|f - \hat{f}_m\|_{L^2}^2,$$

where m^* is the maximal element of \mathcal{M}_n . We compute an approximation of $\|\cdot\|_{L^2}^2$ with Simpson's rule on a bounded interval I of \mathbb{R}_+ with 1000 points. We compute the expectations with an empirical mean over 500 samples (we use the same samples for \mathcal{R}_1 , \mathcal{R}_2 and \mathcal{R}_o).

Remark 4.1. Since f is non-negative, \hat{f}_m is replaced by $(\hat{f}_m)_+ := \max(\hat{f}_m, 0)$ in the following.

Distributions for X . We consider several distributions for X . In each case, we normalize the distribution for the variance to be 1. We use the same examples as Mabon (2017).

- Exponential $\mathcal{E}(1)$, $I = [0, 5]$.
- Gamma $\Gamma(20, \sqrt{20})$, $I = [0, 10]$.
- Rayleigh, $f(x) = \frac{x}{\sigma^2} \exp(-\frac{x^2}{2\sigma^2})$ with $\sigma^2 = \frac{2}{4-\pi}$, $I = [0, 6]$.
- Weibull, $f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$, with $k = \frac{3}{2}$ and $\lambda = 1/\sqrt{\Gamma(1 + \frac{4}{3}) - \Gamma(1 + \frac{2}{3})^2}$, $I = [0, 5]$.
- Beta $B(4, 5)$ with normalisation $\frac{9}{\sqrt{2}}$, $I = [0, 8]$.
- Gamma mixture $0.4\Gamma(2, 2) + 0.6\Gamma(16, 4)$ with normalisation $\frac{1}{\sqrt{2.96}}$, $I = [0, 5]$.

Distributions for Y . We choose gamma distributions for Y . We recall the density of the gamma distribution with parameters $\alpha \geq 1$ and $\lambda > 0$:

$$g(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x \geq 0.$$

with Γ the gamma function. These distributions satisfy assumptions **(A1)** to **(A3)**, so we can use the procedure described in Subsection 3.1. We can compute the Laguerre coefficients of g exactly:

$$\forall k \in \mathbb{N}, \quad b_k = \sqrt{2} \left(\frac{\lambda}{\lambda+1} \right)^\alpha \sum_{j=0}^k \binom{k}{j} \binom{\alpha-1+j}{j} \left(\frac{-2}{1+\lambda} \right)^j$$

where $\binom{\alpha-1+j}{j} := \frac{(\alpha-1+j) \times (\alpha-2+j) \times \dots \times \alpha}{j!}$. Denoting \mathbf{b}_m the vectors $[b_k]_{k \leq m-1}$, we can compute:

$$\mathbf{b}_m = \sqrt{2} \left(\frac{\lambda}{\lambda+1} \right)^\alpha \mathbf{P}_m \mathbf{v}_m$$

where \mathbf{P}_m is the $m \times m$ matrix with components $P_{ij} := \binom{i}{j}$ for $0 \leq i, j \leq m-1$, and \mathbf{v}_m is the vector of size m :

$$\mathbf{v}_m := \left[\binom{\alpha-1+j}{j} \left(\frac{-2}{1+\lambda} \right)^j \right]_{0 \leq j \leq m-1}.$$

Therefore, we can compute easily and efficiently the matrix \mathbf{G}_m and its inverse.

We choose two distributions for Y : the $\Gamma(2, \sqrt{20})$ distribution which has variance $\frac{1}{10}$ and the $\Gamma(2, \sqrt{8})$ distribution which has variance $\frac{1}{4}$.

The constant $c(g)$. This constant in the penalty pen_2 is not known, so we have to evaluate it numerically. Following the proof of the Proposition 2.7, we see this constant appears is the upper bound:

$$\frac{\rho^2(\mathbf{G}_m^{-1})}{n} \sum_{k=0}^{m-1} \mathbb{E}[\varphi_k(Z_1)^2] \leq \frac{c(g) \sqrt{m} \rho^2(\mathbf{G}_m^{-1})}{n},$$

which is true for every density of X_1 and for every m . Therefore,

$$c(g) = \sup_{X_1 \sim f} \sup_{m \in \mathbb{N}^*} \frac{1}{\sqrt{m}} \sum_{k=0}^{m-1} \mathbb{E}[\varphi_k(X_1 + Y_1)^2].$$

For every distribution of X we considered earlier and for every m from 1 to 50, we computed:

$$Q(m) := \frac{1}{\sqrt{m}} \sum_{k=0}^{m-1} \mathbb{E}[\varphi_k(X_1 + Y_1)^2],$$

using an empirical mean over 1000 realizations of (X_1, Y_1) to compute the expectation, then we took the maximum. For the two distributions we considered for Y_1 , the worst constant was about 0.4 and was reached for $X_1 \sim \mathcal{E}(1)$. For the other distributions of X_1 , we found constants between 0.2 and 0.3. To be safe, we chose a bigger constant $c(g) = 0.5$ for both $\Gamma(2, \sqrt{20})$ and $\Gamma(2, \sqrt{8})$ distributions.

The constants κ . We use the same constant $\kappa_1 = 0.03$ as Mabon (2017). Concerning the constant κ_2 , we made several simulations to calibrate it and we chose $\kappa_2 = 0.04$. During the calibration of κ_2 , we realized that the collection \mathcal{M}_n did not have enough models. The selected dimensions were too small and lead to bad performances. Therefore, we considered the following collection:

$$\mathcal{M}'_n := \left\{ 1 \leq m \leq m^* \mid m \rho^2(\mathbf{G}_m^{-1}) \leq 10 \frac{n}{\log n} \right\}.$$

Procedure. We take the maximal dimension to be $m^* = 20$. We use two sample sizes: $n = 200$ and $n = 2000$. Given a distribution for X , a distribution for Y , a sample size n , and a subdivision $\Sigma(I)$ of I , do:

1. Compute the matrix \mathbf{G}_{20} and its inverse.
2. Compute the collection \mathcal{M}'_n .
3. Compute $f(x)$ for $x \in \Sigma(I)$.
4. Repeat 500 times:
 - (a) Generate a sample $Z_i = X_i + Y_i$, $i = 1, \dots, n$.
 - (b) Compute the Laguerre coefficients vector $\hat{\mathbf{c}}_{20}$ and compute $\hat{\mathbf{a}}_{20} = \mathbf{G}_{20}^{-1} \hat{\mathbf{c}}_{20}$.
 - (c) Compute \hat{m}_1 and \hat{m}_2 minimizing:

$$\hat{m}_i = \arg \min_{m \in \mathcal{M}'_n} \left(- \sum_{k=0}^{m-1} \hat{a}_k^2 + \kappa_i \text{pen}_i(m) \right), \quad i \in \{1, 2\},$$

and compute $(\hat{f}_{\hat{m}_1}(x))_+$ and $(\hat{f}_{\hat{m}_2}(x))_+$ for $x \in \Sigma(I)$.

- (d) Compute $J_1 = \|f - (\hat{f}_{\hat{m}_1})_+\|_{L^2}^2$ and $J_2 = \|f - (\hat{f}_{\hat{m}_2})_+\|_{L^2}^2$ using Simpson's rule.
- (e) For m from 1 to 20, compute $(\hat{f}_m(x))_+$ for $x \in \Sigma(I)$; then compute $J(m) = \|f - (\hat{f}_m)_+\|_{L^2}^2$ using Simpson's rule.
5. Compute \mathcal{R}_1 (resp. \mathcal{R}_2) as the mean of J_1 (resp. J_2) over the 500 samples.
6. For each m , compute the mean of $J(m)$ over the 500 samples, then compute \mathcal{R}_o as the minimum of these quantities.

Results. We show our results in Table 1. We note that in the cases $X \sim \mathcal{E}(1)$ and $X \sim \Gamma(20, \sqrt{20})$, both estimators performed badly compared to the oracle. We see that our estimator $\hat{f}_{\hat{m}_2}$ is better when X has gamma, Rayleigh, beta and mixture gamma distribution. The estimator $\hat{f}_{\hat{m}_1}$ is better when X has Weibull and exponential distribution.

For illustration, Figure 1 shows several estimations when the distribution of X is a mixture gamma. It's a bimodal distribution, so it's interesting to see if the estimators are able to recover the two "pikes" and the "hollow" of the true density. For small samples ($n = 200$), the estimator $\hat{f}_{\hat{m}_1}$ seems to fail more often than the estimator $\hat{f}_{\hat{m}_2}$. For large samples ($n = 2000$), both estimators locate well the two pikes (by overestimating the first one and underestimating the second one), but the estimator $\hat{f}_{\hat{m}_1}$ locates less well the hollow and underestimates more the second pike.

4.2. Model selection in dimension 2

In this subsection, our goal is to illustrate the "Goldenshluger and Lepski" procedure on two examples, in the bi-dimensional case $d = 2$. We did not make an extensive study of the estimator as we did in the previous subsection, because our implementation of this procedure is very slow.

| Distribution of X | MISE | $\sigma_Y^2 = 1/10$ | | $\sigma_Y^2 = 1/4$ | |
|-------------------------|-----------------|---------------------|------------|--------------------|------------|
| | | $n = 200$ | $n = 2000$ | $n = 200$ | $n = 2000$ |
| $\mathcal{E}(1)$ | \mathcal{R}_1 | 139 | 8.57 | 113 | 9.95 |
| | \mathcal{R}_2 | 204 | 30.4 | 203 | 37.2 |
| | \mathcal{R}_o | 10.6 | 0.93 | 14.7 | 1.50 |
| $\Gamma(20, \sqrt{20})$ | \mathcal{R}_1 | 127 | 32.1 | 407 | 109 |
| | \mathcal{R}_2 | 80.4 | 16.4 | 350 | 58.0 |
| | \mathcal{R}_o | 26.8 | 3.84 | 34.7 | 5.24 |
| Rayleigh | \mathcal{R}_1 | 68.1 | 7.13 | 88.7 | 8.23 |
| | \mathcal{R}_2 | 42.1 | 6.44 | 50.3 | 9.13 |
| | \mathcal{R}_o | 29.8 | 4.76 | 35.0 | 7.33 |
| Weibull | \mathcal{R}_1 | 55.2 | 4.89 | 79.0 | 6.80 |
| | \mathcal{R}_2 | 64.3 | 8.56 | 70.3 | 10.5 |
| | \mathcal{R}_o | 30.8 | 4.84 | 42.0 | 6.59 |
| B(4, 5) | \mathcal{R}_1 | 122 | 16.8 | 184 | 29.8 |
| | \mathcal{R}_2 | 39.6 | 6.50 | 53.0 | 16.5 |
| | \mathcal{R}_o | 30.9 | 4.52 | 35.5 | 5.59 |
| Γ mixture | \mathcal{R}_1 | 286 | 63.8 | 717 | 82.9 |
| | \mathcal{R}_2 | 162 | 47.1 | 595 | 71.2 |
| | \mathcal{R}_o | 151 | 34.0 | 228 | 49.4 |

Table 1: MISE computation over 500 samples. MISEs are multiplied by 10^4 . The column “ $\sigma_Y^2 = 1/10$ ” corresponds to the case $Y_i \sim \Gamma(2, \sqrt{20})$ and the column “ $\sigma_Y^2 = 1/4$ ” to the case $Y_i \sim \Gamma(2, \sqrt{8})$.

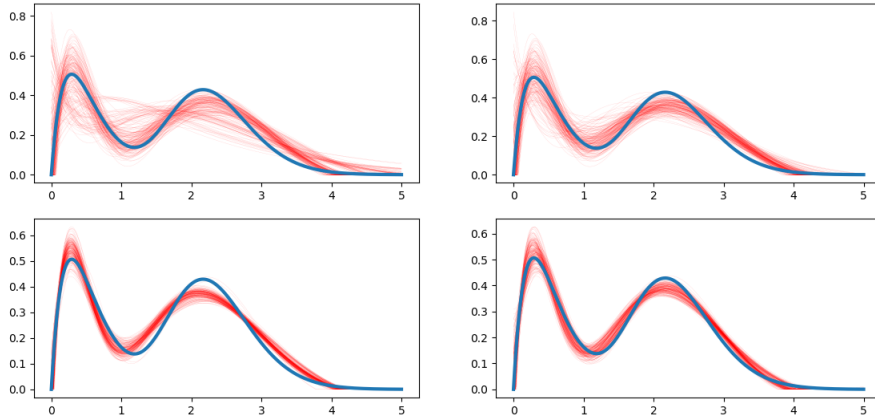


Figure 1: Density estimation when X is a mixture of gamma distributions and when $Y \sim \Gamma(2, \sqrt{20})$. The thick blue curve represents the true density and each thin red curve represents the estimation obtained with one sample (total: 200 samples). First line: $n = 200$, second line: $n = 2000$. Left: $\hat{f}_{\hat{m}_1}$, right: $\hat{f}_{\hat{m}_2}$.

Distributions for X . To generate a random vector $X = (X^{(1)}, X^{(2)})$, we do the following: we generate a random variable $W^{(1)}$ with some distribution and we generate $W^{(2)}$ independently with some other distribution, and we compute:

$$\begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{10} \\ \frac{1}{20} & 1 \end{bmatrix} \begin{bmatrix} W^{(1)} \\ W^{(2)} \end{bmatrix},$$

such that the coordinates of X are not independent. We chose two settings for the distributions of $W^{(1)}$ and $W^{(2)}$:

Case 1. $W^{(1)} \sim \Gamma(3, 1)$, $W^{(2)} \sim \Gamma(2, 1)$;

Case 2. $W^{(1)} \sim \Gamma(3, 1)$, $W^{(2)} \sim \text{B}(4, 5)$ with normalization $\frac{9}{\sqrt{2}}$.

Distribution for Y . We chose distributions of the form $\Gamma(\alpha_1, \lambda_1) \otimes \Gamma(\alpha_2, \lambda_2)$. These distributions satisfy the assumptions **(A1’)** to **(A3’)** of the Subsection 3.2. Moreover, the Laguerre coefficients can be computed easily: if γ_i is the density of the distribution $\Gamma(\alpha_i, \lambda_i)$, then we have $g = \gamma_1 \otimes \gamma_2$, so:

$$\forall k \in \mathbb{N}^2, \quad b_k = \langle \gamma_1, \varphi_{k_1} \rangle_{L^2} \times \langle \gamma_2, \varphi_{k_2} \rangle_{L^2}.$$

Refer to Subsection 4.1 for the computation of $\langle \gamma_i, \varphi_{k_i} \rangle_{L^2}$. We chose the distribution $\Gamma(2, \sqrt{20}) \otimes \Gamma(2, \sqrt{20})$ for Y .

The κ constants. After several simulations, we chose $\kappa_1 = \kappa_2 = 0.0004$. However, this calibration is rough.

Results. We took the maximal model to be $m^* = (13, 13)$. In both cases, we simulated a sample of size $n = 10^4$, then we computed the model \hat{m} . We computed $\|f - \hat{f}_{\hat{m}}\|_{L^2}^2$ by approximating the double integral by a Riemann sum:

$$\iint_{\mathbb{R}_+^2} \left(f(x, y) - \hat{f}_{\hat{m}}(x, y) \right)^2 dx dy \approx \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} \left(f(x_i, y_j) - \hat{f}_{\hat{m}}(x_i, y_j) \right)^2 \Delta x_i \Delta y_j.$$

In case 1, we discretized $[0, 7] \times [0, 5]$ with step size 0.01 for both axes. In case 2, we discretized $[0, 7] \times [0, 6]$ with step size 0.01 for both axes. We did the same with the maximal model estimator $\hat{f}_{(13,13)}$, for comparison.

In case 1, the selected model is $\hat{m} = (4, 3)$, and the L^2 distance between f and the estimation is 1.0×10^{-5} for $\hat{f}_{\hat{m}}$ and 2.0×10^{-5} for $\hat{f}_{(13,13)}$. For illustration, the Figure 2 represents level sets of f , $\hat{f}_{\hat{m}}$ and $\hat{f}_{(13,13)}$.

In case 2, the selected model is $\hat{m} = (3, 8)$, and the L^2 distance between f and the estimators are 1.2×10^{-5} for $\hat{f}_{\hat{m}}$ and 2.4×10^{-5} for $\hat{f}_{(13,13)}$. For illustration, the Figure 3 represents level sets of f , $\hat{f}_{\hat{m}}$ and $\hat{f}_{(13,13)}$.

We emphasize that in both cases, the selected model is anisotropic: the components of \hat{m} are not equal. It is a property we expected from a ‘‘Goldenshluger and Lepski’’ procedure. The estimator adapts to the regularity of f .

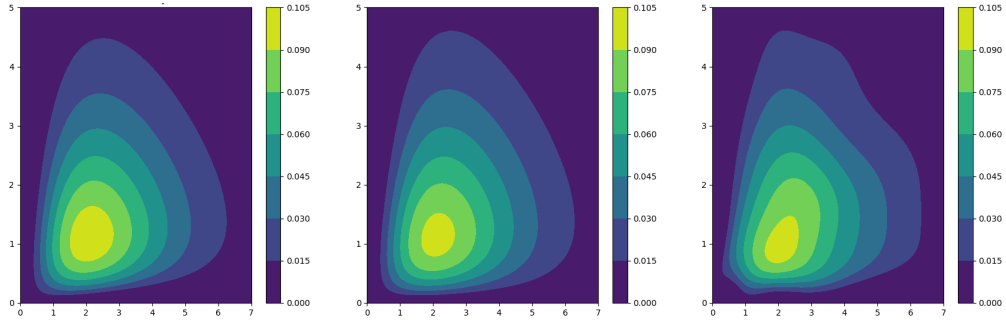


Figure 2: **Case 1**, left: true density, center: estimator $\hat{f}_{\hat{m}}$, right: estimator $\hat{f}_{(13,13)}$.

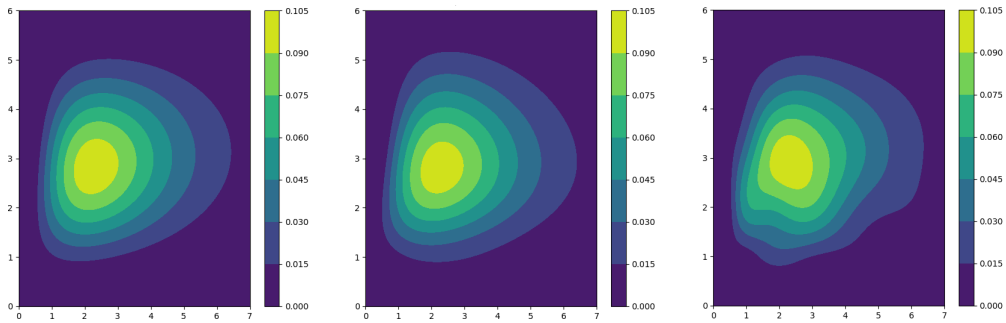


Figure 3: **Case 2**, left: true density, center: estimator $\hat{f}_{\hat{m}}$, right: estimator $\hat{f}_{(13,13)}$.

5. Proofs

5.1. Proofs of Section 2

Proof of Proposition 2.1. Using the relation (6), we have:

$$f * g = \sum_{k \in \mathbb{N}^d} \sum_{j \in \mathbb{N}^d} a_k b_j (\varphi_k * \varphi_j) = \sum_{k \in \mathbb{N}^d} \sum_{j \in \mathbb{N}^d} a_k b_j \bigotimes_{q=1}^d 2^{-1/2} (\varphi_{k_q + j_q} - \varphi_{k_q + j_q + 1}).$$

By setting $\ell_q = k_q + j_q$, we get:

$$f * g = 2^{-d/2} \sum_{\ell \in \mathbb{N}^d} \sum_{\substack{k \in \mathbb{N}^d \\ k \leq \ell}} a_k b_{\ell - k} \bigotimes_{q=1}^d (\varphi_{\ell_q} - \varphi_{\ell_q + 1}) = 2^{-d/2} \sum_{\ell \in \mathbb{N}^d} (a * b)_\ell \bigotimes_{q=1}^d (\varphi_{\ell_q} - \varphi_{\ell_q + 1}).$$

Using the tensor product multi-linearity:

$$\begin{aligned} \bigotimes_{q=1}^d (\varphi_{\ell_q} - \varphi_{\ell_q + 1}) &= \bigotimes_{q=1}^d \sum_{\varepsilon_q=0}^1 (-1)^{\varepsilon_q} \varphi_{\ell_q + \varepsilon_q} = \sum_{\varepsilon_1=0}^1 \cdots \sum_{\varepsilon_d=0}^1 \bigotimes_{q=1}^d (-1)^{\varepsilon_q} \varphi_{\ell_q + \varepsilon_q} \\ &= \sum_{\varepsilon \in \{0,1\}^d} (-1)^{|\varepsilon|} \varphi_{\ell + \varepsilon}. \end{aligned}$$

Thus, we get:

$$f * g = \sum_{\ell \in \mathbb{N}^d} (a * b)_\ell \left[2^{-d/2} \sum_{\varepsilon \in \{0,1\}^d} (-1)^{|\varepsilon|} \varphi_{\ell + \varepsilon} \right] = \sum_{\ell \in \mathbb{N}^d} \left[2^{-d/2} \sum_{\varepsilon \in \{0,1\}^d} (-1)^{|\varepsilon|} (a * b)_{\ell - \varepsilon} \right] \varphi_\ell.$$

Since $h = f * g$, by uniqueness of the Laguerre coefficients of h , we obtain the desired relation. \square

Proof of Proposition 2.3. First, we notice that $\forall \ell$, $\mathbf{G}_{\ell\ell} = \mathbb{E}[e^{-(Y_1^{(1)} + \cdots + Y_1^{(d)})}] > 0$. We proceed by induction on $|k|$.

- If $|k| = 0$, then $k = \mathbf{0}$ and $a_{\mathbf{0}} = (\mathbf{G}_{\mathbf{0},\mathbf{0}})^{-1} c_{\mathbf{0}}$.
- Let $r \in \mathbb{N}$, we suppose (9) is true for every $k \in \mathbb{N}^d$ such that $|k| \leq r$. Let $k \in \mathbb{N}^d$ with $|k| = r + 1$. From (8),

$$c_k = \sum_{\substack{\ell \leq k \\ \ell \neq k}} \mathbf{G}_{k\ell} a_\ell + \mathbf{G}_{kk} a_k.$$

If $\ell \leq k$ with $\ell \neq k$ then $|\ell| < |k|$, so we can use the induction assumption:

$$a_k = (\mathbf{G}_{kk})^{-1} \left(c_k - \sum_{\substack{\ell \leq k \\ \ell \neq k}} \mathbf{G}_{k\ell} \sum_{j \leq \ell} \mathbf{H}_{\ell j} c_j \right).$$

Thus, by setting $\mathbf{H}_{kk} := (\mathbf{G}_{kk})^{-1}$ and $\mathbf{H}_{kj} := (\mathbf{G}_{kk})^{-1} \sum_{j \leq \ell \leq k, \ell \neq k} \mathbf{G}_{k\ell} \mathbf{H}_{\ell j}$ for every $j \leq k$, $j \neq k$, we've just proved (9) for all k such that $|k| = r + 1$. \square

Proof of Proposition 2.4. By Pythagoras theorem, $\|f - \hat{f}_m\|_{\mathbb{L}^2}^2 = \|f - f_m\|_{\mathbb{L}^2}^2 + \|f_m - \hat{f}_m\|_{\mathbb{L}^2}^2$. We decompose the second term on the Laguerre basis:

$$\|f_m - \hat{f}_m\|_{\mathbb{L}^2}^2 = \sum_{k \leq m-1} (a_k - \hat{a}_k)^2 = \|\mathbf{a}_m - \hat{\mathbf{a}}_m\|_{\mathbb{R}^m}^2.$$

We now give an upper bound on the last quantity in two different ways.

1. The first way is a bound using the spectral norm.

$$\begin{aligned} \mathbb{E}\|f_m - \hat{f}_m\|_{\mathbb{L}^2}^2 &= \mathbb{E}\|\mathbf{G}_m^{-1}(\mathbf{c}_m - \hat{\mathbf{c}}_m)\|_{\mathbb{R}^m}^2 \\ &\leq \rho^2(\mathbf{G}_m^{-1}) \mathbb{E}\|\mathbf{c}_m - \hat{\mathbf{c}}_m\|_{\mathbb{R}^m}^2 \\ &= \rho^2(\mathbf{G}_m^{-1}) \mathbb{E}\left[\sum_{k \leq m-1} \left(\frac{1}{n} \sum_{i=1}^n \varphi_k(Z_i) - \mathbb{E}[\varphi_k(Z_1)] \right)^2 \right] \\ &= \frac{\rho^2(\mathbf{G}_m^{-1})}{n} \sum_{k \leq m-1} \text{Var}(\varphi_k(Z_1)) \leq \frac{\rho^2(\mathbf{G}_m^{-1})}{n} \sum_{k \leq m-1} \mathbb{E}[\varphi_k(Z_1)^2]. \end{aligned}$$

Yet for every $x \in \mathbb{R}_+^d$, $\varphi_k(x)^2 \leq 2^d$ and $\#\{k \in \mathbb{N}^d \mid k \leq m-1\} = \prod_{q=1}^d m_q$, so:

$$\mathbb{E}\|f_m - \hat{f}_m\|_{\mathbb{L}^2}^2 \leq \frac{\rho^2(\mathbf{G}_m^{-1})}{n} \sum_{k \leq m-1} \mathbb{E}[\varphi_k(Z_1)^2] \leq \frac{2^d \left(\prod_{q=1}^d m_q \right) \rho^2(\mathbf{G}_m^{-1})}{n}. \quad (17)$$

2. For the second way, we develop the square.

$$\begin{aligned} \mathbb{E}\|\mathbf{G}_m^{-1}(\mathbf{c}_m - \hat{\mathbf{c}}_m)\|_{\mathbb{R}^m}^2 &= \mathbb{E}\left[\sum_{\ell \leq m-1} \left(\sum_{k \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k} (c_k - \hat{c}_k) \right)^2 \right] \\ &= \mathbb{E}\left[\sum_{\ell \leq m-1} \sum_{k \leq m-1} \sum_{k' \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k} (c_k - \hat{c}_k) [\mathbf{G}_m^{-1}]_{\ell k'} (c_{k'} - \hat{c}_{k'}) \right] \\ &= \sum_{\ell \leq m-1} \sum_{k \leq m-1} \sum_{k' \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k} [\mathbf{G}_m^{-1}]_{\ell k'} \text{Cov}(\hat{c}_k, \hat{c}_{k'}) \\ &= \sum_{\ell \leq m-1} \sum_{k \leq m-1} \sum_{k' \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k} [\mathbf{G}_m^{-1}]_{\ell k'} \frac{\text{Cov}(\varphi_k(Z_1), \varphi_{k'}(Z_1))}{n} \\ &= \frac{1}{n} \sum_{\ell \leq m-1} \text{Cov}\left(\sum_{k \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k} \varphi_k(Z_1), \sum_{k' \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k'} \varphi_{k'}(Z_1) \right) \\ &= \frac{1}{n} \sum_{\ell \leq m-1} \text{Var}\left(\sum_{k \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k} \varphi_k(Z_1) \right). \end{aligned}$$

We control the variance by the expectation of the square:

$$\begin{aligned}
\text{Var} \left(\sum_{k \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k} \varphi_k(Z_1) \right) &\leq \mathbb{E} \left[\left(\sum_{k \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k} \varphi_k(Z_1) \right)^2 \right] \\
&= \int_{\mathbb{R}_+^d} \left(\sum_{k \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k} \varphi_k(z) \right)^2 h(z) dz \\
&\leq \|h\|_\infty \left\| \sum_{k \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k} \varphi_k \right\|_{L^2}^2 = \|h\|_\infty \sum_{k \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k}^2.
\end{aligned}$$

In the end, we obtain:

$$\mathbb{E} \|f_m - \hat{f}_m\|_{L^2}^2 \leq \frac{\|h\|_\infty}{n} \sum_{\ell \leq m-1} \sum_{k \leq m-1} [\mathbf{G}_m^{-1}]_{\ell k}^2 = \frac{\|h\|_\infty \|\mathbf{G}_m^{-1}\|_F^2}{n}. \quad \square$$

To prove Proposition 2.7, we will use the following lemma about the order of magnitude of the functions φ_k for $k \in \mathbb{N}$.

Lemma 5.1 (Askey and Wainger 1965). *There exists $k_0 \in \mathbb{N}$ and $C > 0$ such that for every $k \geq k_0$,*

$$\left| \varphi_k \left(\frac{x}{2} \right) \right| \leq C \times \begin{cases} 1 & \text{if } 0 \leq x \leq \frac{1}{k} \\ k^{-\frac{1}{4}} x^{-\frac{1}{4}} & \text{if } \frac{1}{k} \leq x \leq \delta k \\ k^{-\frac{1}{4}} (\nu - x)^{-\frac{1}{4}} & \text{if } \delta k \leq x \leq \nu - \nu^{\frac{1}{3}} \\ k^{-\frac{1}{3}} & \text{if } \nu - \nu^{\frac{1}{3}} \leq x \leq \nu + \nu^{\frac{1}{3}} \\ k^{-\frac{1}{4}} (x - \nu)^{-\frac{1}{4}} \exp(-\eta(x - \nu)^{\frac{3}{2}} \nu^{-\frac{1}{2}}) & \text{if } \nu + \nu^{\frac{1}{3}} \leq x \leq (1 + \lambda)\nu \\ \exp(-\xi x) & \text{if } (1 + \lambda)\nu \leq x \end{cases}$$

where $\nu := 4k + 2$, where δ et λ are small enough positive constants, and where η and ξ are fixed positive constants.

Proof of Proposition 2.7. Following the proof of Proposition 2.4, we see we need to improve the upper bound in (17). Using Lemma 5.1, we have for all $k \in \mathbb{N}$,

$$|\varphi_k(x)|^2 \leq C \begin{cases} 1 & \text{if } 0 \leq 2x \leq \frac{1}{k}, \\ 1/\sqrt{kx} & \text{if } \frac{1}{k} \leq 2x \leq \delta k, \\ R_k & \text{if } 2x \geq \delta k. \end{cases}$$

where $R_k = o(1/\sqrt{k})$ does not depend on x . Since “ $0 \leq x \leq \frac{1}{k}$ ” is equivalent to “ $1 \leq \frac{1}{\sqrt{kx}}$ ”, we get $|\varphi_k(x)|^2 \leq C(\frac{1}{\sqrt{kx}} + R_k)$. Thus for $k = (k_1, \dots, k_d) \in \mathbb{N}^d$, we get:

$$\begin{aligned}
\mathbb{E}[\varphi_k(Z)^2] &= \mathbb{E} \left[\prod_{j=1}^d \varphi_{k_j}(Z^{(j)})^2 \right] \leq \mathbb{E} \left[\prod_{j=1}^d C \left(\frac{1}{\sqrt{k_j Z^{(j)}}} + R_{k_j} \right) \right] \\
&\leq \mathbb{E} \left[\prod_{j=1}^d C \left(\frac{1}{\sqrt{k_j Y^{(j)}}} + R_{k_j} \right) \right] \\
&= \frac{C^d}{\sqrt{k_1 \cdots k_d}} \mathbb{E} \left[\frac{1}{\sqrt{Y^{(1)} \cdots Y^{(d)}}} \right] + \tilde{C} o \left(\frac{1}{\sqrt{k_1 \cdots k_d}} \right)
\end{aligned}$$

where \tilde{C} depends on M_J for $J \subset \{1, \dots, d\}$. Therefore:

$$\sum_{k \leq m-1} \mathbb{E}[\varphi_k(Z)^2] \leq c(g) \sqrt{D_m},$$

with $c(g)$ a positive constant depending on M_J for $J \subset \{1, \dots, d\}$. \square

Proof of Theorem 2.14. Using Proposition 2.4, the bound (12) on the bias term, and Proposition 2.10, if f belongs to $W^s(\mathbb{R}_+^d, L)$, we have:

$$\forall m \in (\mathbb{N}^*)^d, \quad \mathbb{E} \|f - \hat{f}_m\|_{L^2}^2 \leq L \sum_{i=1}^d m_i^{-s_i} + \frac{\|h\|_\infty}{n} \prod_{i=1}^d m_i^{2\alpha_i}.$$

By Remark 2.5, $\|h\|_\infty$ is controlled by $\|f\|_{L^2} \|g\|_{L^2}$. Because $f \in W^s(\mathbb{R}_+^d, L)$, we have $\|f\|_{L^2}^2 \leq L$, so $\|h\|_\infty \leq \sqrt{L} \|g\|_{L^2}$. Thus,

$$\forall m \in (\mathbb{N}^*)^d, \quad \mathbb{E} \|f - \hat{f}_m\|_{L^2}^2 \leq L \sum_{i=1}^d m_i^{-s_i} + \frac{\sqrt{L} \|g\|_{L^2}}{n} \prod_{i=1}^d m_i^{2\alpha_i} =: \psi_n(m).$$

Let m_{opt} minimizing ψ_n , then the gradient of ψ_n vanishes on m_{opt} , so we have:

$$\forall j \in \{1, \dots, d\}, \quad \frac{L s_j}{2\alpha_j} m_{\text{opt},j}^{-s_j} = \frac{\sqrt{L} \|g\|_{L^2}}{n} \prod_{i=1}^d m_{\text{opt},i}^{2\alpha_i}. \quad (18)$$

Therefore, the coordinates of m_{opt} must satisfy $\frac{s_i}{2\alpha_i} m_{\text{opt},i}^{-s_i} = \frac{s_j}{2\alpha_j} m_{\text{opt},j}^{-s_j}$ for every i and j in $\{1, \dots, d\}$. Using these relations in (18), we obtain:

$$\forall j \in \{1, \dots, d\}, \quad m_{\text{opt},j} = C_j(s, L, g) n^{1/(s_j + s_j \sum_{i=1}^d \frac{2\alpha_i}{s_i})}$$

where $C_j(s, L, g)$ are constants depending on s, L and g . The minimum value of ψ_n is then:

$$\psi_n(m_{\text{opt}}) = C(s, L, g) n^{-1/(1 + \sum_{i=1}^d \frac{2\alpha_i}{s_i})}$$

where $C(s, L, g)$ is a constant depending in s, L and g .

Now, if f belongs to $\mathcal{S}^r(\mathbb{R}_+^d, L)$, we use the bound (13) on the bias term:

$$\forall m \in (\mathbb{N}^*)^d, \quad \mathbb{E} \|f - \hat{f}_m\|_{L^2}^2 \leq L \sum_{i=1}^d e^{-r_i m_i} + \frac{\sqrt{L} \|g\|_\infty}{n} \prod_{i=1}^d m_i^{2\alpha_i} =: \phi_n(m).$$

We minimize the function $\phi_n(m)$ as we did above, and we find the following relations:

$$\forall j \in \{1, \dots, d\}, \quad r_j m_{\text{opt},j} e^{-r_j m_{\text{opt},j}} = C_j(r, L, g) \frac{(r_j m_{\text{opt},j})^{\sum_{i=1}^d 2\alpha_j}}{n},$$

where $C_j(r, L, g)$ are constants depending on r, L and g . Taking the log, we find:

$$r_j m_{\text{opt},j} + \left(\sum_{i=1}^d 2\alpha_i - 1 \right) \log(r_j m_{\text{opt},j}) = \log n - \log C_j(r, L, g).$$

Thus, when n goes to $+\infty$, we have $r_j m_{\text{opt},j} \sim \log n$. Taking $m_{\text{opt}} \in (\mathbb{N}^*)^d$ such that $m_{\text{opt},j} \propto \frac{\log n}{r_j}$, we find:

$$\phi_n(m_{\text{opt}}) \leq C(r, L, g) \frac{(\log n)^{\sum_{i=1}^d 2\alpha_i}}{n},$$

where $C(r, L, g)$ is a constant depending on r, L and g . \square

5.2. Proposition 2.10

5.2.1. Preliminary results

To prove this proposition, we first need to extend the theory of Toeplitz matrices to hypermatrices. For more details about Toeplitz matrices, see Böttcher and Grudsky (2005). We say that T is an infinite lower triangular Toeplitz hypermatrix if it is lower triangular according to Definition 2.2 and if the value of $T_{\ell k}$ depends only on the difference $\ell - k$:

$$T = [T_{\ell k}]_{\ell, k \in \mathbb{N}^d}, \quad T_{\ell k} = a_{\ell - k}$$

with $a = [a_k]_{k \in \mathbb{Z}^d}$ and $a_k = 0$ if $k \notin \mathbb{N}^d$. So there is a bijection that takes $a \in \mathbb{R}^{\mathbb{N}^d}$ and returns the corresponding Toeplitz hypermatrix $T(a)$. We can see $T(a)$ as a linear map on $\mathbb{R}^{\mathbb{N}^d}$:

$$x = [x_k]_{k \in \mathbb{N}^d} \mapsto \left[\sum_{k \in \mathbb{N}^d} T_{\ell k} x_k \right]_{\ell \in \mathbb{N}^d} = \left[\sum_{k \leq \ell} a_{\ell - k} x_k \right]_{\ell \in \mathbb{N}^d} = a * x.$$

The associativity of the convolution product gives that for every a and b , we have $T(a) \times T(b) = T(a * b)$.

Notation. If $m \in (\mathbb{N}^*)^d$, we denote by $T_m(a) \in \mathbb{R}^{(m, m)}$ the sub-hypermatrix of $T(a)$ constructed by taking only the coefficients $[T(a)]_{\ell, k}$ for $\ell, k \leq m - \mathbf{1}$. Note that as a linear map on \mathbb{R}^m , we have $T_m(a) = T(a)|_{\mathbb{R}^m}$ because of the triangular structure.

Operator on $\ell^p(\mathbb{N}^d)$. If $a \in \ell^1(\mathbb{N}^d)$ and if x is in $\ell^p(\mathbb{N}^d)$ with $p \in [1, +\infty]$, then $a * x$ belongs to $\ell^p(\mathbb{N}^d)$ we have $\|a * x\|_{\ell^p} \leq \|a\|_{\ell^1} \|x\|_{\ell^p}$. In other words, $T(a)$ is an operator from $\ell^p(\mathbb{N}^d)$ to $\ell^p(\mathbb{N}^d)$ and its operator norm is bounded by $\|a\|_{\ell^1}$.

The ℓ^2 case. We will need to study $T(a)$ as an operator on $\ell^2(\mathbb{N}^d)$ and give a bound on its norm with milder assumptions than $a \in \ell^1(\mathbb{N}^d)$. We define a subspace of $L^2(\mathbb{T}^d)$ that plays an important role in this matter.

Definition 5.2 (Hardy space). We define the Hardy space as the following subset of $L^2(\mathbb{T}^d)$:

$$H^2(\mathbb{T}^d) := \{u \in L^2(\mathbb{T}^d) \mid \forall k \notin \mathbb{N}^d, c_k(u) = 0\}$$

where $c_k(u)$ denotes the k -th Fourier coefficient of u and \mathbb{T} is the set of complex numbers with unitary module.

The map that takes a function and returns its Fourier coefficient is then an isometric bijection between $H^2(\mathbb{T}^d)$ and $\ell^2(\mathbb{N}^d)$.

$$\begin{aligned} \mathcal{F}: H^2(\mathbb{T}^d) &\longrightarrow \ell^2(\mathbb{N}^d) \\ f &\longmapsto [c_k(f)]_{k \in \mathbb{N}^d} \end{aligned}$$

We see that if u and v are $H^2(\mathbb{T}^d)$ functions, the identity $\mathcal{F}[u] * \mathcal{F}[v] = \mathcal{F}[u \times v]$ translates for Toeplitz hypermatrices into $T(\mathcal{F}[u]) \times T(\mathcal{F}[v]) = T(\mathcal{F}[u \times v])$.

Under the additional assumption that $\mathcal{F}^{-1}[a]$ belongs to $L^\infty(\mathbb{T}^d)$, we show in the next proposition that $T(a)$ defines an operator from $\ell^2(\mathbb{N}^d)$ to $\ell^2(\mathbb{N}^d)$.

Proposition 5.3. *Let $a \in \ell^2(\mathbb{N}^d)$ such that $u := \mathcal{F}^{-1}[a] \in L^\infty(\mathbb{T}^d)$. Then $T(a)$ defines an operator on $\ell^2(\mathbb{N}^d)$ and its operator norm is bounded by $\|u\|_\infty$.*

Proof. Let $x \in \ell^2(\mathbb{N}^d)$ and let $v := \mathcal{F}^{-1}[x]$. Then:

$$\|T(a)x\|_{\ell^2} = \|\mathcal{F}[u] * \mathcal{F}[v]\|_{\ell^2} = \|\mathcal{F}[u \times v]\|_{\ell^2} = \|u \times v\|_{L^2} \leq \|u\|_{\infty} \|v\|_{L^2} = \|u\|_{\infty} \|x\|_{\ell^2}. \quad \square$$

Before we can prove Proposition 2.10, some facts need to be established. We denote by $\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ the Riemann sphere. We will use the following functions:

- We make the assumption that β belongs to $\ell^1(\mathbb{N}^d)$ so that the power series $B(z) := \sum_{k \in \mathbb{N}^d} \beta_k z^k$ is normally convergent on $\overline{\mathbb{D}^d}$ and defines a function which is continuous on $\overline{\mathbb{D}^d}$ and holomorphic on \mathbb{D}^d .
- We denote by G the Laplace transform of g . This function is defined on \mathcal{P}_+^d , continuous on \mathcal{P}_+^d , and holomorphic on $\{s \in \mathbb{C}^d \mid \forall q, \Re(s_q) > 0\}$.
- If u is a $L^1(\mathbb{R}_+^d)$ or $L^2(\mathbb{R}_+^d)$ function, we denote by:

$$u^*(\omega) := \int_{\mathbb{R}_+^d} e^{i\omega \cdot x} u(x) dx, \quad \omega \in \mathbb{R}^d,$$

its Fourier transform, where $\omega \cdot x := \sum_i \omega_i x_i$. The Fourier transform of the Laguerre functions φ_k , $k \in \mathbb{N}^d$ can be computed from the case $d = 1$:

$$\varphi_k^*(\omega) = \prod_{q=1}^d \varphi_{k_q}^*(\omega_q) = \prod_{q=1}^d (-1)^{k_q} \sqrt{2} \frac{(1 + i\omega_q)^{k_q}}{(1 - i\omega_q)^{k_q+1}}.$$

We will need to understand the behaviour of the map $z \mapsto \frac{1+z}{1-z}$.

Lemma 5.4. *Let $\eta: \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$ be the homographic function $\eta(s) := \frac{s-1}{s+1}$ (with $\eta(\infty) = 1$ and $\eta(-1) = \infty$).*

1. *The function η is invertible and its inverse is $\eta^{-1}(z) = \frac{1+z}{1-z}$.*
2. *The image of $\{s \in \mathbb{C} \mid \Re(s) > 0\}$ by η is \mathbb{D} .*
3. *The image of $\{s \in \mathbb{C} \mid \Re(s) = 0\}$ by η is $\mathbb{T} \setminus \{1\}$.*

In the next proposition that generalizes lemma C.1 from Comte et al. (2017), we show that the functions G and B are linked through the transformation η .

Proposition 5.5. *If $\theta \in \mathbb{R}^d$, we denote $e^{i\theta} := (e^{i\theta_1}, \dots, e^{i\theta_d})$. If $\beta \in \ell^1(\mathbb{N}^d)$, then $[\beta_k]_{k \in \mathbb{N}^d}$ are the Fourier coefficients of $\theta \mapsto G\left(\frac{1+e^{i\theta}}{1-e^{i\theta}}\right)$ and we have:*

$$\forall \theta \in \mathbb{R}^d, \quad G\left(\frac{1+e^{i\theta}}{1-e^{i\theta}}\right) = \sum_{k \in \mathbb{N}^d} \beta_k e^{ik \cdot \theta} = B(e^{i\theta})$$

with normal convergence of the series.

Remark 5.6. If $g \in W^s(\mathbb{R}_+^d, L)$ with $s \in (1, +\infty)^d$, then β is $\ell^1(\mathbb{N}^d)$, see Remark 2.11.

Proof of Proposition 5.5. We start from the expression of β_k and we use the Plancherel isometry:

$$\begin{aligned} \beta_k &= 2^{-d/2} \sum_{k \leq 1} (-1)^{|\varepsilon|} b_{k-\varepsilon} = 2^{-d/2} \sum_{k \leq 1} (-1)^{|\varepsilon|} \langle g, \varphi_{k-\varepsilon} \rangle_{L^2} \\ &= 2^{-d/2} \sum_{k \leq 1} (-1)^{|\varepsilon|} \frac{1}{(2\pi)^d} \langle g^*, \varphi_{k-\varepsilon}^* \rangle_{L^2} = \left\langle g^*, \frac{2^{-d/2}}{(2\pi)^d} \sum_{\varepsilon \leq 1} (-1)^{|\varepsilon|} \varphi_{k-\varepsilon}^* \right\rangle_{L^2}. \end{aligned}$$

Let us compute the second factor in the scalar product.

$$\begin{aligned}
\frac{2^{-d/2}}{(2\pi)^d} \sum_{\varepsilon \leq \mathbf{1}} (-1)^{|\varepsilon|} \varphi_{k-\varepsilon}^*(\omega) &= \frac{2^{-d/2}}{(2\pi)^d} \sum_{\varepsilon \leq \mathbf{1}} (-1)^{|\varepsilon|} \prod_{q=1}^d (-1)^{k_q - \varepsilon_q} \sqrt{2} \frac{(1 + i\omega_q)^{k_q - \varepsilon_q}}{(1 - i\omega_q)^{k_q - \varepsilon_q + 1}} \\
&= \frac{(-1)^{|k|}}{(2\pi)^d} \sum_{\varepsilon \leq \mathbf{1}} \left(\frac{\mathbf{1} + i\omega}{\mathbf{1} - i\omega} \right)^{k - \varepsilon} \left(\frac{\mathbf{1}}{\mathbf{1} - i\omega} \right)^{\mathbf{1}} \\
&= \left(\frac{i\omega + \mathbf{1}}{i\omega - \mathbf{1}} \right)^k \frac{1}{(2\pi)^d} \left(\frac{\mathbf{1} - i\omega}{\mathbf{1} + i\omega} + \mathbf{1} \right)^{\mathbf{1}} \left(\frac{\mathbf{1}}{\mathbf{1} - i\omega} \right)^{\mathbf{1}} \\
&= \left(\frac{i\omega + \mathbf{1}}{i\omega - \mathbf{1}} \right)^k \frac{1}{\pi^d} \prod_{q=1}^d \frac{1}{1 + \omega_q^2}.
\end{aligned}$$

We derive the following expression:

$$\beta_k = \frac{1}{\pi^d} \int_{\mathbb{R}^d} g^*(\omega) \left(\frac{i\omega - \mathbf{1}}{i\omega + \mathbf{1}} \right)^k \frac{d\omega_1 \cdots d\omega_d}{(1 + \omega_1^2) \cdots (1 + \omega_d^2)}.$$

The change of variable $e^{-i\theta_q} = \frac{i\omega_q - 1}{i\omega_q + 1} = \eta(i\omega_q)$ gives then:

$$\begin{aligned}
\beta_k &= \frac{1}{(2\pi)^d} \int_{[0, 2\pi]^d} g^* \left(\frac{e^{i\theta} + \mathbf{1}}{i(e^{i\theta} - \mathbf{1})} \right) e^{-ik \cdot \theta} d\theta \\
&= \frac{1}{(2\pi)^d} \int_{[0, 2\pi]^d} G \left(\frac{\mathbf{1} + e^{i\theta}}{\mathbf{1} - e^{i\theta}} \right) e^{-ik \cdot \theta} d\theta.
\end{aligned} \tag{19}$$

Therefore, if $k \in \mathbb{N}^d$, the k -th Fourier coefficient of the function $\theta \mapsto G\left(\frac{\mathbf{1} + e^{i\theta}}{\mathbf{1} - e^{i\theta}}\right)$ is β_k . On the other hand, if $k \notin \mathbb{N}^d$, let us show that the Fourier coefficients vanish. Without loss of generality, we assume that $k_1 < 0$, and we compute the k -th Fourier coefficient:

$$\begin{aligned}
&\frac{1}{(2\pi)^d} \int_{[0, 2\pi]^d} G \left(\frac{\mathbf{1} + e^{i\theta}}{\mathbf{1} - e^{i\theta}} \right) e^{-ik \cdot \theta} d\theta \\
&= \frac{1}{\pi^d} \int_{\mathbb{R}^d} g^*(\omega) \left(\frac{i\omega_1 - 1}{i\omega_1 + 1} \right)^{k_1} \prod_{q=2}^d \left(\frac{i\omega_q - 1}{i\omega_q + 1} \right)^{k_q} \frac{d\omega_1 \cdots d\omega_d}{(1 + \omega_1^2) \cdots (1 + \omega_d^2)} \\
&= \frac{1}{\pi^d} \int_{\mathbb{R}^d} g^*(\omega) \left(\frac{i(-\omega_1) - 1}{i(-\omega_1) + 1} \right)^{-k_1} \prod_{q=2}^d \left(\frac{i\omega_q - 1}{i\omega_q + 1} \right)^{k_q} \frac{d\omega_1 \cdots d\omega_d}{(1 + \omega_1^2) \cdots (1 + \omega_d^2)} \\
&= \frac{-1}{\pi^d} \int_{\mathbb{R}^d} g^*(-\omega_1, \omega_2, \dots, \omega_d) \left(\frac{i\omega - \mathbf{1}}{i\omega + \mathbf{1}} \right)^{k'} \frac{d\omega_1 \cdots d\omega_d}{(1 + \omega_1^2) \cdots (1 + \omega_d^2)}
\end{aligned}$$

where $k' := (-k_1, k_2, \dots, k_d)$. Let δ be the map on \mathbb{R}^d defined by $\delta(\omega) := (-\omega_1, \omega_2, \dots, \omega_d)$, then we have:

$$\begin{aligned}
\frac{-1}{\pi^d} \int_{\mathbb{R}^d} g^*(\delta(\omega)) \left(\frac{i\omega - \mathbf{1}}{i\omega + \mathbf{1}} \right)^{k'} \frac{d\omega_1 \cdots d\omega_d}{(1 + \omega_1^2) \cdots (1 + \omega_d^2)} &= \left\langle -g^* \circ \delta, \frac{2^{-d/2}}{(2\pi)^d} \sum_{\varepsilon \leq \mathbf{1}} (-1)^{|\varepsilon|} \varphi_{k'-\varepsilon}^* \right\rangle_{L^2} \\
&= 2^{-d/2} \sum_{k \leq \mathbf{1}} (-1)^{|\varepsilon|} \frac{1}{(2\pi)^d} \langle -g^* \circ \delta, \varphi_{k'-\varepsilon}^* \rangle_{L^2} \\
&= 2^{-d/2} \sum_{k \leq \mathbf{1}} (-1)^{|\varepsilon|} \langle g \circ \delta, \varphi_{k'-\varepsilon} \rangle_{L^2},
\end{aligned}$$

because $(g \circ \delta)^* = -g^* \circ \delta$. This last expression is always zero:

$$\langle g \circ \delta, \varphi_{k' - \varepsilon} \rangle_{L^2} = \int_{\mathbb{R}^d} g(-x_1, x_2, \dots, x_d) \varphi_{k' - \varepsilon}(x) dx = 0$$

since the function g is zero of $x_1 > 0$ and $\varphi_{k' - \varepsilon}$ is zero if $x_1 < 0$. Thus, the Fourier coefficients of $\theta \mapsto G\left(\frac{1+e^{i\theta}}{1-e^{i\theta}}\right)$ are $[\beta_k]_{k \in \mathbb{N}^d}$.

The function G is continuous on $(i\mathbb{R} \cup \{\infty\})^d$, so the function $G \circ (\eta^{-1})^{\otimes d}$ is continuous on \mathbb{T}^d by Lemma 5.4. Therefore, since the Fourier series $\theta \mapsto G\left(\frac{1+e^{i\theta}}{1-e^{i\theta}}\right)$ is normally convergent, this function is equal to its Fourier series at each point:

$$\forall \theta \in \mathbb{R}^d, \quad G\left(\frac{1+e^{i\theta}}{1-e^{i\theta}}\right) = \sum_{k \in \mathbb{N}^d} \beta_k e^{ik \cdot \theta}. \quad \square$$

We need a last technical lemma before we start the proof of Proposition 2.10.

Lemma 5.7. *Let $\alpha \in (\mathbb{N}^*)^d$, then $(\mathbf{1} - z)^{-\alpha}$ admits a power series expansion on \mathbb{D}^d given by:*

$$\forall z \in \mathbb{D}^d, \quad (\mathbf{1} - z)^{-\alpha} = \sum_{j \in \mathbb{N}^d} \binom{\alpha - \mathbf{1} + j}{j} z^j.$$

Moreover, denoting ζ_j the j -th coefficient in the power series above, for $m \geq 4$ we have $\|T_m(\zeta)\|_F^2 \leq m^{2\alpha}$ where $T(\zeta)$ is the Toeplitz hypermatrix constructed from the coefficients ζ_j .

Proof of Lemma 5.7. We recall the following identity: for $z \in \mathbb{D}$ and $r \in \mathbb{N}^*$, we have $(1-z)^{-r} = \sum_{j=0}^{\infty} \binom{r-1+j}{j} z^j$. Thus, for $z \in \mathbb{D}^d$ and $\alpha \in (\mathbb{N}^*)^d$, we have:

$$(\mathbf{1} - z)^{-\alpha} = \prod_{q=1}^d (1 - z_q)^{-\alpha_q} = \prod_{q=1}^d \sum_{j_q=0}^{\infty} \binom{\alpha_q - 1 + j_q}{j_q} z_q^{j_q} = \sum_{j \in \mathbb{N}^d} \binom{\alpha - \mathbf{1} + j}{j} z^j.$$

Therefore,

$$\begin{aligned} \|T_m(\zeta)\|_F^2 &= \sum_{j \leq m - \mathbf{1}} \binom{\alpha - \mathbf{1} + j}{j}^2 \#\{(k, \ell) \in \mathbb{N}^d \times \mathbb{N}^d \mid k \leq \ell \leq m - \mathbf{1}, \ell - k = j\} \\ &= \prod_{q=1}^d \sum_{j_q=0}^{m_q - 1} \binom{\alpha - 1 + j_q}{j_q}^2 (m_q - j_q), \end{aligned}$$

and it reduces to the case $d = 1$ which was already solved in (Comte et al., 2017, appendix C). So if $m_q \geq 4$ for every q , then $\|T_m(\zeta)\|_F^2 \leq \prod_{q=1}^d m_q^{2\alpha_q} = m^{2\alpha}$. \square

5.2.2. Proof of Proposition 2.10

From Proposition 5.5, we get $G = B \circ \eta^{\otimes d}$. We define a function w on $(\overline{\mathbb{D}} \setminus \{1\})^d$ by:

$$\forall z \in (\overline{\mathbb{D}} \setminus \{1\})^d, \quad w(z) := (\mathbf{1} - z)^{-\alpha} B(z).$$

This function is related to K_α by the identity:

$$\forall z \in (\overline{\mathbb{D}} \setminus \{1\})^d, \quad w(z) = \mathbf{2}^{-\alpha} K_\alpha((\eta^{-1})^{\otimes d}(z)).$$

Thus, the function w can be extended as a function on $\overline{\mathbb{D}}^d$, still denoted w , and according to our assumptions on K_α and Lemma 5.4, w satisfies:

- $w|_{\mathbb{T}^d}$ is continuous;
- w is continuous on $(\overline{\mathbb{D}} \setminus \{1\})^d$;
- w is holomorphic on \mathbb{D}^d ;
- w doesn't vanish on $\overline{\mathbb{D}}^d$.

Thus, the function $w^{-1} = 1/w$ is well defined on $\overline{\mathbb{D}}^d$ and has the same properties. In particular, since it is holomorphic on \mathbb{D}^d , it admits a power series expansion:

$$\forall z \in \mathbb{D}^d, \quad w^{-1}(z) = \sum_{k \in \mathbb{N}^d} d_k z^k.$$

Let us have look on what is happening on \mathbb{T}^d . For $r \in [0, 1[$, we denote w_r^{-1} the function on \mathbb{T}^d defined by:

$$\forall t \in \mathbb{T}^d, \quad w_r^{-1} = w^{-1}(rt) = \sum_{k \in \mathbb{N}^d} d_k r^{|k|} t^k.$$

On the one hand, the Fourier coefficients of w_r^{-1} are $d_k r^{|k|}$ (we set $d_k = 0$ if $k \notin \mathbb{N}^d$). On the other hand, we can compute the Fourier coefficients and we get:

$$\forall k \in \mathbb{Z}^d, \quad d_k r^{|k|} = \frac{1}{(2\pi)^d} \int_{[0, 2\pi]^d} w^{-1}(r e^{i\theta}) e^{-ik \cdot \theta} d\theta.$$

Since w^{-1} is continuous on $(\overline{\mathbb{D}} \setminus \{1\})^d$, we have $w_r^{-1} \rightarrow w|_{\mathbb{T}^d}^{-1}$ a.e. on \mathbb{T}^d . By dominated convergence, we obtain:

$$\forall k \in \mathbb{Z}^d, \quad d_k = \frac{1}{(2\pi)^d} \int_{[0, 2\pi]^d} w^{-1}(e^{i\theta}) e^{-ik \cdot \theta} d\theta.$$

Therefore, (d_k) are the Fourier coefficients of $w|_{\mathbb{T}^d}^{-1}$. Thus, we have shown that $w|_{\mathbb{T}^d}^{-1} \in H^2(\mathbb{T}^d)$ and:

$$w^{-1}(t) = \sum_{k \in \mathbb{N}^d} d_k t^k,$$

with $L^2(\mathbb{T}^d)$ -convergence of the series.

By Lemma 5.7, the function $(\mathbf{1}-z)^{-\alpha}$ admits a power series expansion on \mathbb{D}^d given by $(\mathbf{1}-z)^{-\alpha} = \sum_{k \in \mathbb{N}^d} \zeta_k z^k$. Thus, the power series equality $B(z) \times (\mathbf{1}-z)^{-\alpha} \times w^{-1}(z) = 1$ on the domain \mathbb{D}^d translates to their coefficients into the relation $\beta * \zeta * d = \delta_{\mathbf{0}}$, where $\delta_{\mathbf{0}}$ is the element of $\mathbb{R}^{\mathbb{N}^d}$ defined by:

$$[\delta_{\mathbf{0}}]_k = \begin{cases} 1 & \text{if } k = \mathbf{0} \\ 0 & \text{else.} \end{cases}$$

Taking the corresponding Toeplitz hypermatrices, we get $\mathbf{G} \times T(\zeta) \times T(d) = \mathbf{I}_{\mathbb{N}^d}$, where $\mathbf{I}_{\mathbb{N}^d}$ is the infinite hypermatrix given by:

$$[\mathbf{I}_{\mathbb{N}^d}]_{\ell k} = \begin{cases} 1 & \text{if } \ell = k \\ 0 & \text{else.} \end{cases}$$

Thus, for $m \in \mathbb{N}^*$, we get $\mathbf{G}_m^{-1} = T_m(\zeta) \times T_m(d)$. Taking the Frobenius norm, we obtain the following inequality:

$$\|\mathbf{G}_m^{-1}\|_F^2 \leq \|T_m(\zeta)\|_F^2 \times \rho^2(T_m(d)).$$

From Proposition 5.3, we have:

$$\rho^2(T_m(d)) = \sup_{x \in \mathbb{R}^m \setminus \{0\}} \frac{\|T_m(d)x\|_{\mathbb{R}^m}^2}{\|x\|_{\mathbb{R}^m}^2} \leq \sup_{x \in \ell^2(\mathbb{N}) \setminus \{0\}} \frac{\|T(d)x\|_{\ell^2}^2}{\|x\|_{\ell^2}^2} = \|w_{|\mathbb{T}^d}^{-1}\|_{\infty}^2,$$

and by Lemma 5.7, we have $\|T_m(\zeta)\|_F^2 \leq m^{2\alpha}$ if $m \geq 4$. \square

5.3. Proofs of Section 3

Proof of Theorem 3.2. The proof is identical to the one in Mabon (2017), but we use our Lemma 3.1 instead of Mabon's proposition 7.1. \square

Proof of Lemma 3.1. We will prove that $\|\hat{f}_m - f_m\|_{L^2}^2$ can be written as the supremum of an empirical process, then we will use Talagrand's inequality (Theorem Appendix A.1). Let $m \in (\mathbb{N}^*)^d$, for every $t \in S_m$ we set:

$$\mathbf{G}_m^{-1}(t) := \sum_{k \leq m-1} [\mathbf{G}_m^{-1} \times_d \mathbf{t}_m]_k \varphi_k,$$

where $\mathbf{t}_m \in \mathbb{R}^m$ is the matrix of Laguerre coefficients of t . We define the empirical process ν_n on the space S_{m^*} by:

$$\nu_n(t) := \frac{1}{n} \sum_{i=1}^n (\mathbf{G}_{m^*}^{-1}(t)(Z_i) - \mathbb{E}[\mathbf{G}_{m^*}^{-1}(t)(Z_i)]).$$

Note that S_{m^*} contains every S_m as a subspace for $m \in \mathcal{M}_n$ and because of the triangular structure of \mathbf{G}^{-1} , we have for every $t \in S_m$, $\mathbf{G}_{m^*}^{-1}(t) = \mathbf{G}_m^{-1}(t)$. For $m \in \mathcal{M}_n$, we have then:

$$\|\hat{f}_m - f_m\|_{L^2}^2 = \sum_{k \leq m-1} (\hat{a}_k - a_k)^2 = \sum_{k \leq m-1} \nu_n^2(\varphi_k).$$

The map $t \mapsto \nu_n(t)$ being linear on S_m and $(\varphi_k)_{k \leq m-1}$ being a basis of S_m , we have by Cauchy-Schwarz inequality:

$$\sum_{k \leq m-1} \nu_n^2(\varphi_k) = \sup_{t \in B_m} \nu_n^2(t), \text{ where } B_m := \{t \in S_m \mid \|t\|_{L^2} = 1\}.$$

Thus, we have:

$$\|\hat{f}_m - f_m\|_{L^2}^2 = \sup_{t \in B_m} \nu_n^2(t).$$

We apply Talagrand's inequality¹. We need the constants M, H and v to apply this inequality.

- Computation of M :

$$\begin{aligned} \sup_{t \in B_m} \sup_{x \in \mathbb{R}_+^d} |\mathbf{G}_m^{-1}(t)(x)| &= \sup_{t \in B_m} \sup_{x \in \mathbb{R}_+^d} \left| \sum_{k \leq m-1} [\mathbf{G}_m^{-1} \times_d \mathbf{t}_m]_k \varphi_k(x) \right| \\ &\leq 2^{d/2} \sup_{t \in B_m} \sum_{k \leq m-1} |[\mathbf{G}_m^{-1} \times_d \mathbf{t}_m]_k| \\ &\leq 2^{d/2} \sqrt{D_m} \sup_{t \in B_m} \|\mathbf{G}_m^{-1} \times_d \mathbf{t}_m\|_{\mathbb{R}^m} = 2^{d/2} \sqrt{D_m} \rho(\mathbf{G}_m^{-1}). \end{aligned}$$

Thus, $M^2 := 2^d D_m \rho^2(\mathbf{G}_m^{-1})$.

¹this inequality concerns countable families of functions, but it's not a problem here since ν_n is continuous on S_{m^*} and B_m is separable.

- Computation of H :

$$\begin{aligned} \mathbb{E} \left[\sup_{t \in B_m} |\nu_n(t)| \right]^2 &= \mathbb{E} \left[\|\hat{f}_m - f_m\|_{L^2} \right]^2 \leq \mathbb{E} \left[\|\hat{f}_m - f_m\|_{L^2}^2 \right] \\ &\leq \frac{c(g)\sqrt{D_m} \rho^2(\mathbf{G}_m^{-1})}{n} \wedge \frac{\|g\|_\infty \|\mathbf{G}_m^{-1}\|_F^2}{n}, \end{aligned}$$

by Proposition 2.7. Thus, we take:

$$H^2 := \frac{c(g)\sqrt{D_m} \rho^2(\mathbf{G}_m^{-1})}{n} \wedge \frac{(\|g\|_\infty \vee 1) \|\mathbf{G}_m^{-1}\|_F^2}{n}.$$

- Computation of v :

$$\begin{aligned} \sup_{t \in B_m} \text{Var}(\mathbf{G}_m^{-1}(t)(Z_1)) &\leq \sup_{t \in B_m} \mathbb{E}[\mathbf{G}_m^{-1}(t)(Z_1)^2] = \sup_{t \in B_m} \int_{\mathbb{R}_+^d} \mathbf{G}_m^{-1}(t)(x)^2 h(x) dx \\ &\leq \|h\|_\infty \sup_{t \in B_m} \|\mathbf{G}_m^{-1}(t)\|_{L^2}^2 \\ &\leq \|g\|_\infty \sup_{t \in B_m} \|\mathbf{G}_m^{-1} \times_d \mathbf{t}_m\|_{\mathbb{R}^m}^2 = \|g\|_\infty \rho^2(\mathbf{G}_m^{-1}). \end{aligned}$$

Thus, $v := \|g\|_\infty \rho^2(\mathbf{G}_m^{-1})$.

We consider two cases.

1. “ $c(g)\sqrt{D_m} \rho^2(\mathbf{G}_m^{-1}) \leq (\|g\|_\infty \vee 1) \|\mathbf{G}_m^{-1}\|_F^2$ ” case, then we have:

$$\frac{nH^2}{v} = \frac{c(g)}{\|g\|_\infty} \sqrt{D_m}, \quad \frac{nH}{M} = \frac{\sqrt{c(g)}}{2^{d/2}} \sqrt{n} D_m^{-1/4} \geq \frac{\sqrt{c(g)}}{2^{d/2}} n^{1/4},$$

since $D_m \leq n$. For $\delta > 0$, Talagrand’s inequality gives:

$$\begin{aligned} \mathbb{E} \left[\left(\|\hat{f}_m - f_m\|_{L^2}^2 - 2(1+2\delta)H^2 \right)_+ \right] &\lesssim \frac{\|g\|_\infty \rho^2(\mathbf{G}_m^{-1})}{n} \exp \left(-K\delta \frac{c(g)}{\|g\|_\infty} \sqrt{D_m} \right) \\ &\quad + \frac{D_m \rho^2(\mathbf{G}_m^{-1})}{C(\delta)^2 n^2} \exp \left(-\frac{KC(\delta)\sqrt{2\delta}}{7} \frac{\sqrt{c(g)}}{2^{d/2}} n^{1/4} \right). \end{aligned}$$

We control the first term using assumption (A3’):

$$\sum_{m \in \mathcal{M}_n} \frac{\|g\|_\infty \rho^2(\mathbf{G}_m^{-1})}{n} \exp \left(-K\delta \frac{c(g)}{\|g\|_\infty} \sqrt{D_m} \right) \leq \frac{\|g\|_\infty K(\delta)}{n},$$

where $K(\delta)$ is a constant not depending on n . To control the second term, we use that $D_m \rho^2(\mathbf{G}_m^{-1}) \leq n$ when $m \in \mathcal{M}_n$:

$$\sum_{m \in \mathcal{M}_n} \frac{D_m \rho^2(\mathbf{G}_m^{-1})}{C(\delta)^2 n^2} \exp \left(-\frac{KC(\delta)\sqrt{2\delta}}{7} \frac{\sqrt{c(g)}}{2^{d/2}} n^{1/4} \right) \lesssim \exp \left(-\tilde{K}(\delta) n^{1/4} \right) = o \left(\frac{1}{n} \right),$$

so we have our result.

2. “ $c(g)\sqrt{D_m}\rho^2(\mathbf{G}_m^{-1}) \geq (\|g\|_\infty \vee 1)\|\mathbf{G}_m^{-1}\|_F^2$ ” case, then we have:

$$\frac{nH^2}{v} = \frac{\|\mathbf{G}_m^{-1}\|_F^2}{\rho^2(\mathbf{G}_m^{-1})} \geq 1, \quad \frac{nH}{M} = \frac{\sqrt{n}(\|g\|_\infty \vee 1)^{1/2}\|\mathbf{G}_m^{-1}\|_F}{2^{d/2}\sqrt{D_m}\rho(\mathbf{G}_m^{-1})} \geq \frac{\sqrt{n}}{2^{d/2}}D_m^{-1/2}.$$

Let us show that $\rho(\mathbf{G}_m^{-1}) \geq 1$. We notice that $\frac{2^{d/2}}{b_0}$ is an eigenvalue of \mathbf{G}_m^{-1} : indeed, if E_0 is the elementary hypermatrix with 1 in position $(0, \dots, 0)$ and zeros elsewhere, then:

$$\mathbf{G}_m^{-1} \times_d E_0 = \frac{2^{d/2}}{b_0} E_0.$$

Since the operator norm is bigger than the spectral radius, we have:

$$\rho(\mathbf{G}_m^{-1}) \geq \frac{2^{d/2}}{b_0} = \frac{1}{\mathbb{E}\left[e^{-(Y_1^{(1)} + \dots + Y_1^{(d)})}\right]} \geq 1.$$

Thus, if $m \in \mathcal{M}_n$ we have $D_m \leq D_m \rho^2(\mathbf{G}_m^{-1}) \leq \frac{n}{\log n}$. Therefore, $\frac{nH}{M} \geq \frac{\sqrt{\log n}}{2^{d/2}}$. We apply Talagrand’s inequality with $\delta = a \log n$ and $a > 0$ to be chosen later:

$$\mathbb{E}\left[\left(\|\hat{f}_m - f_m\|_{L^2}^2 - 2(1 + 2a \log n)H^2\right)_+\right] \lesssim \frac{\|g\|_\infty \rho^2(\mathbf{G}_m^{-1})}{n} \frac{1}{n^{Ka}} + \frac{D_m \rho^2(\mathbf{G}_m^{-1})}{C(a \log n)^2 n^2} \exp\left(-\frac{KC(a \log n)\sqrt{2a} \log n}{7 \cdot 2^{d/2}}\right).$$

We assume that² $n \geq \exp(3/a)$, such that $C(a \log n) = 1$. Then:

$$\mathbb{E}\left[\left(\|\hat{f}_m - f_m\|_{L^2}^2 - 2(1 + 2a \log n)H^2\right)_+\right] \lesssim \frac{\|g\|_\infty \rho^2(\mathbf{G}_m^{-1})}{n^{1+Ka}} + \frac{D_m \rho^2(\mathbf{G}_m^{-1})}{n^2} \exp\left(-\frac{K\sqrt{2a}}{7 \times 2^{d/2}} \log n\right).$$

We control the first term using that if $m \in \mathcal{M}_n$ then $\rho^2(\mathbf{G}_m^{-1}) \leq n$:

$$\sum_{m \in \mathcal{M}_n} \frac{\|g\|_\infty \rho^2(\mathbf{G}_m^{-1})}{n^{1+Ka}} \leq \frac{\|g\|_\infty}{n^{Ka-1}} \leq \frac{\|g\|_\infty}{n}$$

if $a \geq 2/K = 12$. To control the second term, we use that if $m \in \mathcal{M}_n$ then $D_m \rho^2(\mathbf{G}_m^{-1}) \leq n$:

$$\sum_{m \in \mathcal{M}_n} \frac{D_m \rho^2(\mathbf{G}_m^{-1})}{n^2} \exp\left(-\frac{K\sqrt{2a}}{7 \times 2^{d/2}} \log n\right) \leq n^{-\frac{K\sqrt{2a}}{7 \times 2^{d/2}}} \leq \frac{1}{n}$$

if $a \geq \frac{2^d \times 7^2}{2K^2}$. □

Proof of Theorem 3.3. Let $m \in \mathcal{M}_n$,

$$\|\hat{f}_{\hat{m}} - f\|_{L^2}^2 \leq 3\|\hat{f}_{\hat{m}} - \hat{f}_{m \wedge \hat{m}}\|_{L^2}^2 + 3\|\hat{f}_{m \wedge \hat{m}} - \hat{f}_m\|_{L^2}^2 + 3\|\hat{f}_m - f\|_{L^2}^2. \quad (20)$$

²actually, we assume later that $a \geq 12$, so there is no assumptions on n after all.

By definition of $A(m)$ and \hat{m} , and since $\kappa_1 \leq \kappa_2$, we get:

$$\begin{aligned} \|\hat{f}_{\hat{m}} - \hat{f}_{m \wedge \hat{m}}\|_{\mathbb{L}^2}^2 + \|\hat{f}_{m \wedge \hat{m}} - \hat{f}_m\|_{\mathbb{L}^2}^2 &\leq A(m) + \kappa_1 V(\hat{m}) + A(\hat{m}) + \kappa_1 V(m) \\ &\leq 2A(m) + (\kappa_1 + \kappa_2)V(m). \end{aligned}$$

Thus, by taking the expectation in (20) and using Proposition 2.7, we obtain:

$$\mathbb{E}\left[\|\hat{f}_{\hat{m}} - f\|_{\mathbb{L}^2}^2\right] \leq 6\mathbb{E}[A(m)] + 3(\kappa_1 + \kappa_2)V(m) + 3(\|f - f_m\|_{\mathbb{L}^2}^2 + V(m)). \quad (21)$$

Now, we need to control $\mathbb{E}[A(m)]$. Let $m' \in \mathcal{M}_n$.

$$\|\hat{f}_{m'} - \hat{f}_{m \wedge m'}\|_{\mathbb{L}^2}^2 \leq 3\|\hat{f}_{m'} - f_{m'}\|_{\mathbb{L}^2}^2 + 3\|f_{m'} - f_{m \wedge m'}\|_{\mathbb{L}^2}^2 + 3\|f_{m \wedge m'} - \hat{f}_{m \wedge m'}\|_{\mathbb{L}^2}^2.$$

Thus,

$$\begin{aligned} A(m) &\leq 3 \left[\max_{m' \in \mathcal{M}_n} \left(\|\hat{f}_{m'} - f_{m'}\|_{\mathbb{L}^2}^2 - \frac{\kappa_1}{6} V(m') \right)_+ \right. \\ &\quad \left. + \max_{m' \in \mathcal{M}_n} \left(\|f_{m \wedge m'} - \hat{f}_{m \wedge m'}\|_{\mathbb{L}^2}^2 - \frac{\kappa_1}{6} V(m') \right)_+ + \max_{m' \in \mathcal{M}_n} \|f_{m'} - f_{m \wedge m'}\|_{\mathbb{L}^2}^2 \right]. \end{aligned} \quad (22)$$

We denote these three terms T_1 , $T_2(m)$ and $T_3(m)$, in that order.

Bound on $\mathbb{E}[T_1]$. We control the maximum by the sum on all terms:

$$T_1 \leq \sum_{m' \in \mathcal{M}_n} \left(\|\hat{f}_{m'} - f_{m'}\|_{\mathbb{L}^2}^2 - \frac{\kappa_1}{6} V(m') \right)_+.$$

By Lemma 3.1, there exist κ_0 such that for every choice of $\kappa_1 > \kappa_0$, $\mathbb{E}[T_1] \leq C/n$ for some constant C depending on κ_1 and g .

Bound on $\mathbb{E}[T_2(m)]$. We decompose the maximum by looking at which indices j we have $m'_j < m_j$:

$$\begin{aligned} T_2(m) &= \max_{J \subseteq \{1, \dots, d\}} \max_{\substack{m' \in \mathcal{M}_n \\ \{j \mid m'_j < m_j\} = J}} \left(\|f_{m \wedge m'} - \hat{f}_{m \wedge m'}\|_{\mathbb{L}^2}^2 - \frac{\kappa_1}{6} V(m') \right)_+ \\ &\leq \max_{J \subseteq \{1, \dots, d\}} \max_{\substack{m' \in \mathcal{M}_n \\ \{j \mid m'_j < m_j\} = J}} \left(\|f_{m \wedge m'} - \hat{f}_{m \wedge m'}\|_{\mathbb{L}^2}^2 - \frac{\kappa_1}{6} V(m \wedge m') \right)_+, \end{aligned} \quad (23)$$

since V is non-decreasing in its arguments. For $J \subseteq \{1, \dots, d\}$ fixed, we introduce the collection:

$$\mathcal{M}_n(m, J) := \{m' \in \mathcal{M}_n \mid \forall j \in J, m'_j < m_j \text{ and } \forall j \notin J, m'_j = m_j\}.$$

Then we notice that $\{j \mid m'_j < m_j\} = J$ if and only if $m \wedge m' \in \mathcal{M}_n(m, J)$. So (23) becomes:

$$\begin{aligned} T_2(m) &\leq \max_{J \subseteq \{1, \dots, d\}} \max_{m' \in \mathcal{M}_n(m, J)} \left(\|f_{m'} - \hat{f}_{m'}\|_{\mathbb{L}^2}^2 - \frac{\kappa_1}{6} V(m') \right)_+ \\ &\leq 2^d \sum_{m' \in \mathcal{M}_n} \left(\|f_{m'} - \hat{f}_{m'}\|_{\mathbb{L}^2}^2 - \frac{\kappa_1}{6} V(m') \right)_+. \end{aligned}$$

Then, as we did with T_1 , we have for every choice of $\kappa_1 > \kappa_0$, $\mathbb{E}[T_2(m)] \leq 2^d \times C/n$.

Bound on $T_3(m)$. For $m' \in \mathcal{M}_n$ fixed, let $J_{m,m'} := \{j \mid m_j < m'_j\}$ and let $S_{m,m'}$ be the following closed subspace of $L^2(\mathbb{R}_+^d)$:

$$S_{m,m'} := \overline{\text{Span}}\{\varphi_k : k \in \mathbb{N}^d \text{ and } \forall j \in J_{m,m'}, k_j \leq m_j - 1\}.$$

Then we notice that $S_{m'} \cap S_{m,m'} = S_{m \wedge m'}$ and that $S_m \subseteq S_{m,m'}$. We denote by $\Pi_{m'}$ the projection on $S_{m'}$ and $\Pi_{m,m'}$ the projection on $S_{m,m'}$. Then, because $\Pi_{m'}$ is 1-lipschitz,

$$\|f_{m'} - f_{m \wedge m'}\|_{L^2}^2 = \|\Pi_{m'}(f - \Pi_{m,m'}f)\|_{L^2}^2 \leq \|f - \Pi_{m,m'}f\|_{L^2}^2 \leq \|f - f_m\|_{L^2}^2,$$

where the last inequality comes from the definition of the projection on $S_{m,m'}$. So $T_3(m)$ is controlled by the bias term $\|f - f_m\|_{L^2}^2$.

End of the proof. We use the bounds on $\mathbb{E}[T_1]$, $\mathbb{E}[T_2(m)]$ and $T_3(m)$ to control $\mathbb{E}[A(m)]$ in (21):

$$\mathbb{E}\left[\|f - \hat{f}_{\hat{m}}\|_{L^2}^2\right] \leq 9\|f - f_m\|_{L^2}^2 + 3(1 + \kappa_1 + \kappa_2)V(m) + \frac{6(1 + 2^d)C}{n}.$$

Since this bound holds for every $m \in \mathcal{M}_n$, we get:

$$\mathbb{E}\left[\|f - \hat{f}_{\hat{m}}\|_{L^2}^2\right] \leq \max\{9, 3(1 + \kappa_1 + \kappa_2)\} \inf_{m \in \mathcal{M}_n} (\|f - f_m\|_{L^2}^2 + V(m)) + \frac{6(1 + 2^d)C}{n}. \quad \square$$

Acknowledgments

The author wants to thank Fabienne Comte and Céline Duval for their helpful advice.

Funding

This work was supported by a grant from Région Île-de-France.

Appendix A. Talagrand's inequality

This concentration inequality is derived from the work of Klein and Rio (2005). A proof can be found in the PhD thesis of Chagny (2013a, chapter 2).

Theorem Appendix A.1 (Talagrand's inequality). *Let $n \in \mathbb{N}^*$ and let \mathcal{F} be a set at most countable of measurable functions on a polish space E . Let ξ_1, \dots, ξ_n be independent random variables with values in E . We consider ν_n the centred empirical process defined by:*

$$\nu_n(f) := \frac{1}{n} \sum_{i=1}^n (f(\xi_i) - \mathbb{E}[f(\xi_i)]), \quad f \in \mathcal{F}.$$

We assume that there exists three positive constants M , H and v such that:

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M, \quad \mathbb{E}\left[\sup_{f \in \mathcal{F}} |\nu_n(f)|\right] \leq H, \quad \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \text{Var}(f(\xi_i)) \leq v.$$

Then for every $\delta > 0$, we have the following inequality:

$$\mathbb{E}\left[\left(\sup_{f \in \mathcal{F}} \nu_n^2(f) - 2(1 + 2\delta)H^2\right)_+\right] \leq \frac{4}{K} \left(\frac{v}{n} e^{-K\delta \frac{nH^2}{v}} + \frac{49M^2}{K C(\delta)^2 n^2} e^{-\frac{KC(\delta)\sqrt{2\delta}}{7} \frac{nH}{M}}\right),$$

with $C(\delta) := (\sqrt{1 + \delta} - 1) \wedge 1$ and $K = 1/6$.

References

- Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Number 55 in Applied Mathematics Series. National Bureau of Standards, New York, NY, 10th edition, 1972.
- Richard Askey and Stephen Wainger. Mean Convergence of Expansions in Laguerre and Hermite Series. *American Journal of Mathematics*, 87(3):695, July 1965.
- Bruno Bongioanni and José L. Torrea. What is a Sobolev space for the Laguerre function systems? *Studia Mathematica*, 192(2):147–172, 2009.
- Cristina Butucea. Deconvolution of supersmooth densities with smooth noise. *Canadian Journal of Statistics*, 32(2):181–192, June 2004.
- Cristina Butucea and Alexandre B. Tsybakov. Sharp Optimality in Density Deconvolution with Dominating Bias. I. *Theory of Probability & Its Applications*, 52(1):24–39, January 2008a.
- Cristina Butucea and Alexandre B. Tsybakov. Sharp Optimality in Density Deconvolution with Dominating Bias. II. *Theory of Probability & Its Applications*, 52(2):237–249, January 2008b.
- Albrecht Böttcher and Sergei M. Grudsky. *Spectral properties of banded Toeplitz matrices*. Society for Industrial and Applied Mathematics, 2005.
- Gaëlle Chagny. *Estimation adaptative avec des données transformées ou incomplètes. Application à des modèles de survie*. PhD thesis, Université Paris Descartes, 2013a.
- Gaëlle Chagny. Warped bases for conditional density estimation. *Mathematical Methods of Statistics*, 22(4):253–282, October 2013b.
- Fabienne Comte and Valentine Genon-Catalot. Adaptive Laguerre density estimation for mixed Poisson models. *Electronic Journal of Statistics*, 9(1):1113–1149, 2015.
- Fabienne Comte and Valentine Genon-Catalot. Laguerre and Hermite bases for inverse problems. *Journal of the Korean Statistical Society*, 47(3):273–296, September 2018.
- Fabienne Comte and Claire Lacour. Anisotropic adaptive kernel deconvolution. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 49(2):569–609, May 2013.
- Fabienne Comte, Yves Rozenholc, and Marie-Luce Taupin. Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics*, 34(3):431–452, September 2006.
- Fabienne Comte, Charles-A. Cuenod, Marianna Pensky, and Yves Rozenholc. Laplace deconvolution on the basis of time domain data and its application to dynamic contrast-enhanced imaging. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1): 69–94, January 2017.
- Aurore Delaigle and Irène Gijbels. Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics*, 56(1):19–47, March 2004.
- Jianqing Fan. On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems. *The Annals of Statistics*, 19(3), September 1991.

- Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608–1632, June 2011.
- Martin L. Hazelton and Berwin A. Turlach. Nonparametric density deconvolution by weighted kernel estimators. *Statistics and Computing*, 19(3):217–228, September 2009.
- Martin L. Hazelton and Berwin A. Turlach. Semiparametric Density Deconvolution. *Scandinavian Journal of Statistics*, 37(1):91–108, 2010.
- Leslie Hogben. *Handbook of Linear Algebra*, chapter 15. Chapman and Hall/CRC, 2nd edition, 2013.
- Thierry Klein and Emmanuel Rio. Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, 33(3):1060–1077, May 2005.
- Oleg Lepski and Thomas Willer. Oracle inequalities and adaptive estimation in the convolution structure density model. *The Annals of Statistics*, 47(1):233–287, February 2019.
- Gwennaëlle Mabon. Adaptive Deconvolution on the Non-negative Real Line. *Scandinavian Journal of Statistics*, 44(3):707–740, September 2017.
- Elias Masry. Multivariate probability density deconvolution for stationary random processes. *IEEE Transactions on Information Theory*, 37(4):1105–1115, July 1991.
- Marianna Pensky and Brani Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, 27(6), December 1999.
- Gilles Rebelles. Structural adaptive deconvolution under Lp-losses. *Mathematical Methods of Statistics*, 25(1):26–53, January 2016.
- Ousmane B. Sacko. Hermite density deconvolution. *Preprint hal-01978591*, 2019.
- Leonard A. Stefanski and Raymond J. Carroll. Deconvolving kernel density estimators. *Statistics*, 21(2):169–184, January 1990.
- Élie Youndjé and Martin T. Wells. Optimal bandwidth selection for multivariate kernel deconvolution density estimation. *TEST*, 17(1):138–162, May 2008.